Open Access & Open Data at PUSH Universities

JUNE 2018

AUTHORS:
ANNE MIMS ADRIAN
BRYTNI EMISON
RUTHIE MUSKER
JESSICA HOPKINSON

IN PARTNERSHIP WITH:
Open Access & Open Data at PUSH Universities
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Foreword</td>
</tr>
<tr>
<td>4</td>
<td>Report Overview</td>
</tr>
<tr>
<td>6</td>
<td>Introduction</td>
</tr>
<tr>
<td>9</td>
<td>Defining Open Access and Open Data</td>
</tr>
<tr>
<td>10</td>
<td>Methodology</td>
</tr>
<tr>
<td>12</td>
<td>Findings</td>
</tr>
<tr>
<td>23</td>
<td>GODAN Donor Open Data Policy and Practice Report</td>
</tr>
<tr>
<td>25</td>
<td>Recommendations to Advance Open Data</td>
</tr>
<tr>
<td>28</td>
<td>Appendix 1: Interview Questions</td>
</tr>
<tr>
<td>29</td>
<td>Appendix 2: Terminology</td>
</tr>
<tr>
<td>30</td>
<td>Citations</td>
</tr>
</tbody>
</table>
Governments, foundations, and the private sector invest billions of dollars each year in agriculture and nutrition research for food security, and much of this research is conducted in universities worldwide. With widespread hunger, malnutrition, and climate uncertainty across the globe, the research that universities perform is essential for solving some of the world’s most urgent concerns.

In a field as dynamic and diverse as agriculture and nutrition, efficient data-driven decisions can contribute to sustainable livelihoods and drive progress toward the Sustainable Development Goal of “zero hunger”. From research design to publication, university faculty and students are involved in all steps along the data value chain across a wide range of fundamental and applied disciplines, ranging from agronomy, animal husbandry and natural resources to human nutrition, food safety and economics, all of which are related to ensuring food insecurity. Researchers collect, analyze, and reuse data, and the new knowledge they create informs important decisions made by governments and industry, while university librarians organize and curate data in large repositories.

Unfortunately, a significant part of this data is lost, misplaced, or locked in closed repositories kept within the walls of the universities. If, however, all universities across the globe developed open data policies and released their data, think of how much more efficiently decisions could be made.

We understand that creating a culture of open data in universities is challenging. This is where PUSH can lead the way. The 99 universities whose presidents have pledged to unite in this global fight against malnutrition and hunger are in a unique position to show by example, to demonstrate that yes, it can be done, and it does save lives and can improve everyone’s well-being across the world.

We hope this report will help you think about how data created through university research fits into the global data ecosystem and how university policies and research community culture need to change to promote data sharing. If you are a researcher, we hope this report will encourage you to incorporate data management plans into your research design. If you are a university policymaker, we hope you will engage with faculty to create open data policies and encourage your colleagues to do so.

We would like to thank PUSH for its tremendous efforts in completing this report. PUSH has been an essential and appreciated GODAN partner in both this report and in our common pledge for open data in agriculture and nutrition.

Together, we can and we will set the data free.

ANDRÉ LAPERRIÈRE
GODAN Executive Director

CATHERINE WOTEKI, PHD
Former USDA Under Secretary for Research, Education and Economics
RATIONALE
The sharing of research findings, as well as other data, is believed to increase the pace of innovation, research breakthroughs, and collaborative problem-solving. Often, however, these data are not readily available, visible, or accessible, resulting in needlessly duplicated research or critical gaps in information. This has led many public research funders (e.g., NSF, NIH, USAID), as well as private donors (e.g., Gates Foundation), to require public universities and other higher education institutions to develop or enhance data management plans that allow for open access and data sharing. While creating a culture with policies and infrastructure platforms that allows for open access and open data is a challenge, it is a challenge that is becoming increasingly necessary for universities to address.

PURPOSE OF STUDY
This exploratory study was conducted by the Hunger Solutions Institute at Auburn University on behalf of Presidents United to Solve Hunger (PUSH), in partnership with the Global Open Data for Agriculture and Nutrition (GODAN) initiative. The purpose of the project was to provide a snapshot of where the 99 universities in the PUSH network were in developing, using, and promoting open access and open data policies and practices by identifying: issues and obstacles, level of support provided to faculty, useful resources and best practices, and comparison of findings to the GODAN donor report.

DEFINITIONS
Although often used interchangeably, the following definitions of open access and open data were used for this project:
- Open access - the free online availability of research articles coupled with the rights to use these articles fully in the digital environment
- Open data - research data, freely available on the internet, that anyone can use or share to download, copy, analyze, re-process, pass to software, or use for any other purpose

METHODOLOGY
The protocol for conducting the study included:
- Website review of the 99 PUSH universities to identify online information/policies about intellectual property, data ownership, open access, and open data
- In-depth, online video interviews with personnel from nine PUSH universities who had advanced knowledge about their university’s open access and open data policies and practices, repositories, level of university support, etc.
- Comparison of findings to GODAN, Donor Open Data Policy and Practice report
FINDINGS
Only 15 of the 99 PUSH university websites provided open access for content, articles, presentations, etc. Of the nine universities participating in the interviews, no single university model emerged as a comprehensive approach or best practice for data management policy, infrastructure and administrative support. However, interviewees did indicate a general agreement, supported by the literature, that open data should be: accessible, machine-readable, high quality, continuously updated, possess unique identifiers, readily linked to other data sources, and contain an open license to reuse the data with credit to the original source.

SAMPLING OF INTERVIEWEES’ RESPONSES TO OPEN ACCESS AND OPEN DATA SURVEY
Benefits
- Provides a public good
- Enhances and accelerates research and innovation
- Increases transparency
- Increases citations and recognition of faculty and their universities
- Increases potential of identifying research collaborators
- Meets funding requirements

Barriers
- Privacy of research subjects
- Data security
- Concerns re stolen work or misuse of data
- Funders’ contract requirements
- Protection of intellectual property
- Lack of resources
- Lack of understanding regarding licensing options and ownership of data

Challenges
- Conflicting attitudes of ownership
- Culture of protectionism
- Pressures linked to ideals of serving the public good
- Lack of technical expertise in sharing data responsibly
- Minimal or no understanding of alternative licenses
- Dearth of resources to support the needed infrastructure to support and maintain data
- Alignment of expectations and requirements between funders and universities

RECOMMENDATIONS TO ADVANCE OPEN DATA
- Effectively communicate the benefits of open access and open data
- Create alignment between funders’ expectations and universities’ capabilities
- Improve faculty’s ability to comply with open access and open data requirements
- Refine and/or develop open access and open data standards and protocols
- Create sufficient infrastructure to handle data planning and management
Introduction

Sharing research findings, as well as data in general, is believed to increase the pace of innovation, research, and collaborative problem-solving. Yet, because data are not readily available, visible, or accessible, research is needlessly duplicated or critical information gaps are left unfilled. Many research funders, such as the National Science Foundation (NSF), National Institutes of Health (NIH), Department for International Development (DFID), United States Agency for International Development (USAID), and the Bill and Melinda Gates Foundation (BMGF) may require open data management plans that include sharing research data. However, university-based research grantees are not always equipped with the resources and skills to manage open data properly and to comply with their funders’ requirements.

The Association of American Universities (AAU) and the Association of Public & Land-Grant Universities (APLU) worked together to create the Public Access Working Group Report and Recommendations. This report emphasizes the inevitability and necessity of open data and the need for change and development in funders’ and universities’ policies. It also highlights barriers and recommendations to open access and open data.

With similar aims, GODAN commissioned the current work reported in this document. The U.S. Government, through its support of the GODAN Secretariat, requested that this project be part of the 2017/2018 GODAN Secretariat work plan. The study, which fits within the context of GODAN’s current research and policy projects, is consistent with its belief that open data can hasten and enhance research and innovation, addressing issues of food and nutrition security.

PUSH, a GODAN partner, was responsible for conducting the investigation. PUSH is a consortium of universities from around the world focused on inspiring collective action to end hunger and malnutrition, both locally and globally. As discoverers, analyzers, curators, and disseminators of data, universities can have a central role in advancing the open data revolution. To date, however, there has been no way to easily assess open access and open data policies, support, protocols, and practices. Nor have there been easy ways to identify workable practices that strike a balance between sharing research data responsibly with
regard to discoverability and utility while maintaining needed privacy and security. Each university and research institution is, therefore, navigating this difficult challenge on its own without the benefit of others’ experiences, lessons learned, and best practices.

The overall purpose of this project was to provide a snapshot of where PUSH universities are in developing, using, and promoting open access and open data. The project was further designed to identify issues and obstacles that universities confront in developing policies, procedures, and platforms for open access and open data. The study explored how universities are supporting faculty to meet open data and data management requirements of funders, as well as identifying resources and best practices utilized at universities that are launching open access and open data processes. Finally, this project provided the opportunity to compare its findings with the GODAN report, “Donor Open Data Policy and Practice: An Analysis of Five Agriculture Programmes”¹.
Open access and open data are terms that are sometimes used interchangeably. Certain definitions may resonate more or less with various audiences and research foci. For the purpose of this project the following definitions were adopted.

### OPEN ACCESS AND OPEN DATA

#### OPEN ACCESS

- **Definition:** the free online, availability of research articles coupled with the rights to use these articles fully in the digital environment.
  - usually the lowest “tier” of open data

#### OPEN DATA

- **Definition:** research data that anyone can access, use or share:
  - freely available on the internet permitting any user to
    - download
    - copy
    - analyze
    - re-process
    - pass to software
    - use for any other purpose
  - NO financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself

Regardless of how the terms are defined, specifically, open data must be released responsibly with clear ownership and licensing. The FAIR Principles, developed by the Dutch TechCentre for the Life Sciences (DTL) specified that data should be Findable, Accessible, Interoperable, and Re-usable to enhance the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.

The types of open data held in the highest regard in quality and utility, which were confirmed by the interviewees, contain the following characteristics:

- Accessible to those who need it
- Machine-readable
- High-quality
- Continuously updated
- Possess unique identifiers
- Able to be linked to other data sources
- Have an open license to reuse the data in any way as long as the original source is credited
Methodology

This study was exploratory in nature and not intended to be highly scientific. Instead, its purpose was to provide PUSH and GODAN with general trends and perspectives that can be addressed in progressing toward greater sharing of research data that enhances discoverability and usefulness. Starting May 2017, the PUSH open data team adopted the following protocol in conducting the research:

**Determined which PUSH universities had open access and open data policies.** The research team reviewed the websites of the 99 PUSH universities, searching for information about intellectual property, ownership, open access, and open data, to ascertain which universities had policies in place.

**Identified which university representatives had the advanced knowledge of institutional policy pertaining to open access and open data to interview.** The choice of universities to approach was based on those that had been actively involved in PUSH with the goal of covering a cross section of institutions including large and small, public and private, and historically black universities. Eighteen PUSH Points of Contact, representing their university presidents/chancellors, identified the interviewees as those at their respective universities who had the most knowledge about open access and open data policies and practices, resulting in interviews of representatives from nine universities (see Table 1).

**Conducted online video interviews to understand the scope and focus of any policies governing open access and open data at selected universities, as well as why policies were not in place at others.** Interviews included discussions about policies, open data practices, and repositories within the universities. They also addressed standards, support for open access, open data, and data management. Appendix 1 outlines the questions that guided the interviews.

**Compared findings with the GODAN report on Donor Open Data Policy and Practice.** Due to the interdependencies between universities and donors, it was important to compare progress in the movement toward open data and to derive ways that the two communities can learn from each other’s policies in determining a viable path forward.

**Concluded with recommendations for universities to develop or improve open access and open data policies.** Recommendations were based on findings generated through the online policy review and interviews.

Due to its exploratory method, useful lessons were learned, not only about information gleaned from the online reviews and interviews, but also from the process of collecting the data. Only 15 out of the 99 university
websites revealed any content related to open access and open data. While representatives of 18 universities were contacted for interviews, only half of that number responded. The final data collection involved online video conversations with 13 female and seven male respondents at nine universities. Despite concerted efforts to collect information from a diverse set of institutions, the final sample of schools included six land-grant, one private, one university system, and one non-U.S. (see Table 1).

Chances of completing interviews were better where PUSH Points were actively invested and involved in the process. At several universities there seemed to be a lack of understanding and/or appreciation of open data, as well as considerable confusion about who was responsible or who “owned” the open access and open data domain. Clearly, university personnel are very busy and sometimes choose not to participate in studies or respond to requests for information. Such was the case here in several instances, making call-backs sometimes nonexistent and interviews very difficult to schedule and complete.

**TABLE 1. PUSH UNIVERSITIES/REPRESENTATIVES INTERVIEWED**

<table>
<thead>
<tr>
<th>UNIVERSITY</th>
<th>Open Access Policies</th>
<th>IT Directors /Tech Transfer Specialists</th>
<th>Librarians /Data Management Specialists</th>
<th>Sponsored Program Associates</th>
<th>VP/Associate Deans for Research</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auburn University</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Cornell University</td>
<td>X</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michigan State University</td>
<td>X</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mississippi State University</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Stenden University</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Tufts University</td>
<td>X</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>University of California</td>
<td>X</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>University of Missouri</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Utah State University</td>
<td>X</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>5</td>
<td>3</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>13</td>
<td>7</td>
</tr>
</tbody>
</table>
Findings

OPEN ACCESS AND OPEN DATA POLICIES
Fifteen of the 99 PUSH universities were found to have open access policies, primarily to meet licensing requirements for sharing scholarly work. One particular open access policy of note at the University of California (UC) specified that “research articles authored by faculty at all 10 UC campuses will be made available to the public at no charge”\textsuperscript{11}. UC faculty recognized the benefits of open access to include greater recognition for themselves and their institution; increased citations; more thorough review, consideration, and critique of their work; and a general increase in scientific, scholarly, and critical knowledge.

At other universities where open access policies were in place, unless the funder mandated it, the policy requirement sometimes tended to be waived. The cost of providing open access was clearly a concern at a number of institutions. Some universities budget for academic journal open access services that range from $2,000 to $5,000. However, other universities leave faculty to find funds on their own, sometimes provided for in departmental budgets. Among the universities that did not have open access policies, several were considering launching those discussions.

During the interviews, many respondents expressed that, philosophically, open access and open data are best for advancing research and could be seen as the responsibility of public institutions working toward the greater good. However, universities are finding themselves in a “paradox of intellectual property” where there is a growing belief in the benefits of open data, yet ideas and new knowledge are often protected for securing patents and as sources of profitization.

While no university in this study explicitly mandated open data as a requirement, most were discussing the topic as it relates to complying with funders’ open data and data management requirements. Only one university was actively looking at developing an open data policy, and the cost issue emerged once again. It was the contention of librarians at several institutions that if an open data policy were in place, there would not be enough resources to support faculty and to maintain data repositories.


text box

\textit{Interviewees were adamant that the development of open data policies should be initiated and promoted by the faculty and supported by the university.}\n
Interviewees were adamant that the development of open data policies should be initiated and promoted by the faculty
and supported by the university. In other words, an open data policy should be developed by faculty and endorsed at all university administrative levels. However, this policy change should be implemented without imposing over-reaching operational demands on researchers.

BENEFITS AND BARRIERS TO OPEN ACCESS AND OPEN DATA

The driving forces in research data being uploaded and shared openly are perceived to be 1) funders’ requirements, and 2) The White House Office of Science and Technology Policy (OSTP) executive order and memo of 2013\(^1\). When interviewees were asked about open data as a practice, they described benefits of open access and open data as:

- Provides a public good
- Enhances and accelerates research and innovation
- Provides more transparency
- Increases citations and recognition for the faculty and their universities
- Increases potential of finding research collaborators
- Meets funding requirements

Interviewees also acknowledged that advances in technologies make it easier to share research and data and the discussions focused on the concerns and roadblocks to embracing open data:

- Privacy of research subjects
- Security of data
- Fears of stolen work or misuse of data
- Funders’ contract requirements
- Protection of university and faculty intellectual property
- Resources needed to provide functional and secure data management
- Ownership of research—published and not published
- Faculty not understanding licensing options

No single model emerged as a comprehensive approach or best practice, encompassing policy, infrastructure, and support for sharing and supporting open data.

INTELLECTUAL PROPERTY AND DATA OWNERSHIP AMBIGUITY

Open data and data management are often believed to be aligned with university Intellectual Property (IP), though data ownership is not explicitly outlined in universities’ policies. Universities have policies that define who owns products developed by faculty and often provide services to manage the IP of those products through technology transfer offices. IP guidelines define what products should be protected and when work can be shared. The de facto modes of practice, not explicitly written into policy, but assumed, is that the faculty members decide how they can make use of the data.
For some universities, faculty’s works are owned by both the university and the faculty. The assumption is that faculty members can decide data’s use, unless dictated by contracts. In some cases, the universities own the copyright to all published work, but policies do not explicitly say who owns the data when the data have not been published. Open data policies would force universities to specifically say who owns the data.

Though faculty are encouraged to license data to clarify intentions of how they want their data used through open Creative Commons licenses, some believe that data could be classified as facts which cannot be licensed.

UNFUNDED MANDATES
Open data as a requirement is often seen as an unfunded mandate. There are costs to providing infrastructure, support, and long-term maintenance of research data. In some cases, with coordination of the library or other support, budgeting for the process of sharing and managing data properly can be built into the proposal. While funders do not prohibit budgeting for open data, some of the costs for uploading and maintaining open data occur outside of the project period.

INFRASTRUCTURE, SYSTEMS, AND REPOSITORIES
All PUSH universities’ websites were searched for open access and open data portals and repositories. Sixty-eight (68) of the 99 PUSH universities provide open access to dissertations, theses, research reports, and outreach articles. A few universities had open access repositories for dissertations, theses, and other student work only. Some universities use open access portals to also share research data. The following PUSH universities have open data repositories: Cornell University eCommons; University of Illinois Databank; Pennsylvania State University Scholarsphere and Datacommons; and University of California System (10 universities) Dash.

Open Data Repositories
Many universities are taking “a wait and see” approach to developing institutional data repositories. Rapidly changing technology and costly resource requirements were often cited as reasons to wait. One librarian said, “When creating a repository, you need to go big or use other repositories.” Developing and maintaining functional repositories that have quality data, have adequate storage for growth, and are secure, free, accessible, interoperable, and reusable require many resources now and in the future.

Most librarians interviewed routinely recommend researchers to use subject-specific repositories instead of institutional repositories. Many subject-specific repositories provide better opportunities for the data to be discovered by others in their fields, and the universities do not have the costly responsibility for infrastructure and maintenance. This is especially true in the case of big
data sets, such as those in medical or health-related fields, where libraries do not have the requisite funding support. When subject-specific repositories are not available, librarians often encourage sharing articles and data on third party sites, like Figshare. Sites like OpenDOAR can be used to identify appropriate repositories.

Advantages of third party hosting services and subject-specific repositories include:
- Assignment of unique digital object identifiers (DOI) to data sets and publications (e.g., articles or presentations)
- Enhanced discoverability within the research community
- Greater integration with publishers’ platforms, author submissions, and peer review systems

Non-Institutional Repositories
Two interviewees referenced the Inter-university Consortium for Political and Social Research (ICPSR) as a well-known, well developed, curated, and often used repository for social sciences. ICPSR offers support, different levels of access, and is free for users, including up to 2 GB of data sharing. However, curation of the data (including assigning a DOI, creating metadata, and a review of data) is usually a one-time fee, varying from $3,000-$8,000 (A. Pienta, personal communication, January 19 2018).

Funding agencies and organizations, such as the NIH, want to make sure that data generated by their grantees are safe, uncorrupted, and shareable. Two associate deans for research noted commercial medical, pharmaceutical, and veterinary medicine data warehouses as extraordinarily valuable resources that meet the highest quality standards. While these data warehouses are not open, are subject to strict privacy regulation, and are fee-based (pay-to-play), they have the capacity to store and secure big data sets, such as those in the field of genomics. As an example, subject-specific databases, such as National Center for Biotechnology Information (NCBI), are available for big data and have policies in place for publishing massive data sets, including entry into a federal database.

An additional feature of many of these databases, as well as a requirement for funding, is the sharing of researchers’ profiles. According to an interviewee, some funders mandate that all investigators and grant applicants maintain a publicly accessible bibliography. The added value is the profiles linked to the publications, thus enabling faculty to identify potential research collaborators.

SUPPORT FROM TECHNOLOGY TRANSFER AND SPONSORED PROGRAMS OFFICES
As previously mentioned, most faculty are sharing their research data because of funding mandates which require data management plans. Sharing data
openly is not enough. Data management, standard formats, protocols, licensing, detailed descriptions, and metadata are all important to make research discoverable and usable. There was a recognition that most faculty need support and assistance from their institutions to adhere to these standards and best practices. Better coordination between sponsored programs offices and libraries would enhance compliance and create greater success in proposal acceptance. Consulting with technology transfer and sponsored programs offices may be needed for faculty to understand when and how data can be shared. When data management is an afterthought in proposals, there is an undue burden on the library to help make effective data plans.

At two universities, the sponsored programs offices were reported to take an active role, making sure that data management plans are addressed in grant proposals. Utah State University's policy encourages researchers to coordinate with the sponsored programs office and the library to ensure that data management plans will meet the funders' logical requirements. This process, which includes creating formal records for the work and data, helps to identify appropriate repositories and often becomes an educational exercise for the researcher.

Most faculty are open and receptive to librarians helping them understand the process, including metadata strategy, description of the data, repository identification, and data management. Many libraries also conduct educational seminars and workshops on the benefits of open access, open data, and data management planning. Since data management is rapidly becoming an integral part of the research process, a couple of interviewees even suggested that data management planning become a required part of the course of study for graduate students.

Most universities have only one data management specialist who works across disciplines. Several universities, however, are planning to hire additional data management specialists to meet the demands of open data requirements and data management. Library support and faculty awareness of that support varies across institutions. In a few cases, when universities had distributive support models, practices and support of open data were also pushed down to the college level. The following list is typical of the support that university libraries provide for open data and data management plans:
- Identifying which repository is best for the research, including an assessment of the longevity of the platform
- Using appropriate open and nonproprietary file formats, such as JSON, XML, RDF, CSV
- Packaging data with metadata, research methods, and protocols for others to use the data without contacting them
- Writing data management plans that will be accepted by the funders
- Coding such as JSON and XML when needed
- Describing the data to include supporting documentation
- Ensuring that data sets are citable
- Ensuring that each data set has a DOI
- Specifying data licensing and reuse rights
- Identifying and prioritizing data for release
- Considering and making adjustments for privacy and ethical considerations
- Using data publishing standards
- Stating concrete commitments to supporting the data and engaging with re-users
- Agreeing and practicing open policy and policy maintenance

STANDARDS AND PROTOCOLS
When asked about standards and protocols, most librarians interviewed discussed using specific nonproprietary formats and metadata descriptions. A few librarians said they depend on DUBLIN Core, a metadata scheme that can be subject specific for describing digital and physical resources. Librarians also told us that they strongly encourage the use of nonproprietary formats that are machine readable. Regardless of preferred schema, data sets must have detailed, standardized descriptions of methodologies and variables that can be discovered through internet searches which can be replicated. In sum, the current practices of the interviewees seem to focus on making sure that others can use the data to confirm the results, reuse the data, and make sense of the data.

Although the interviewers expected to hear more about protocols and standards in identifying the data and using particular ontologies, only one person talked about naming protocols for metadata and harmonizing data to easily link with other data. Some respondents mentioned that they were helping faculty upload data, develop metadata, and use subject specific ontologies. They further indicated that linking to other data would come later with more mature systems and more advanced knowledge and skills of faculty. Most librarians work with faculty on small and medium-sized data sets and refer faculty to subject specific repositories for large data sets.

The lack of discussion on interoperability may have been because the librarians were helping faculty only with small to medium-sized data sets. In some sciences which have well-established subject specific repositories have their own standards and protocols, faculty work directly with those organizations and don’t depend on university librarians. Other than using common formats, like XML or RDF, standards and protocols are not consistent across disciplines. Naming protocols and harmonizing data will be important to make the most of sites, like CGIAR Big Data Platforms, successful.
DEMONSTRATED SUCCESS OF OPEN DATA
A few interviewees cited that both the readership and dissemination of the faculty’s work are increasing as the result of publications and data being shared openly. Because research that is more discoverable leads to faculty who are more discoverable, new collaborations are developing based on a scholar’s body of work rather than just a single publication. In addition, greater faculty recognition results in greater recognition and an enhanced reputation for the university. While some interviewees mentioned the increased citing of work as a plus, another warned that “we may be experiencing some growing pains trying to shoehorn the metrics of success for open data into the traditional promotion and tenure process”.

While most of those interviewed could not think of any singular, concrete example of where an innovative idea was developed from openly shared data at their universities, they did note that there were examples of openly shared data in biological and medicinal research. Two research deans described how advances in research are being enhanced because data are shared and curated within large databases. These databases usually reside on commercial sites that require membership. Interviewees also observed that data repositories were not available or as widely used in other fields, yet they believe in the potential benefit of solving complex problems, such as hunger, through the use of shared data.

After conducting further research, the PUSH open data team found the following models which have demonstrated success in various areas and aspects from which the agricultural field might benefit.
<table>
<thead>
<tr>
<th>SUCCESS</th>
<th>MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development of interoperable standards and protocols</td>
<td>CGIAR Big Data Platform demonstrates the increasing interoperability of the data by using standards and protocols in agriculture research.</td>
</tr>
<tr>
<td>Increased access</td>
<td>Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA-Ref) allows participants access to Jupyter Notebook and RStudio environments to analyze data.</td>
</tr>
<tr>
<td>Professional collaborations</td>
<td>Data Intensive Landscape Limnology Lab at Michigan State University involves multi-disciplinary collaborations to research freshwater ecosystems on a macroscale.</td>
</tr>
<tr>
<td>Professional and student collaborations</td>
<td>European Council for Nuclear Research (CERN) allows students and the general public to participate in coding activities and other methods of analyzing large data sets.</td>
</tr>
<tr>
<td>Community collaborations</td>
<td>The Midwest Big Data Hub (MBDH) collects and manages data from communities concerning natural and built environments, healthcare, and biomedical research. This data is use to improve health services, transportation, and agriculture around the Midwest.</td>
</tr>
<tr>
<td>Educating through open data analysis</td>
<td>The Technical University of Dortmund is the first to create a Master’s level course using the data from CERN to teach students about analyzing data. Additionally CERN has a variety of resources devoted to engaging students of all ages in data analysis and coding.</td>
</tr>
</tbody>
</table>
CHALLENGES OF OPEN DATA
Despite benefits to universities and researchers, the challenges of making research data accessible, open, reusable, and interoperable are complicated. Issues include:
- Conflicting attitudes of ownership
- Culture of protectionism
- Pressures linked to ideals of serving the public good
- Minimal or no understanding of alternative licenses
- Lack of technical expertise in sharing data responsibly
- Dearth of resources to support the infrastructure needed to store and maintain the data

Table 3 provides a list of challenges to creating a culture of open data that were described by the interviewees.

TABLE 3. CHALLENGES TO CREATING A CULTURE OF OPEN DATA

<table>
<thead>
<tr>
<th>ISSUE</th>
<th>THE CHALLENGE</th>
<th>COUNTER PERSPECTIVES OR RESOLUTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-compliance</td>
<td>Open access compliance was low (22%) for physical science research funding from Canadian Institutes of Health Research (CIHR) in 8 years of implementation of its open access policy.</td>
<td>A stronger method of enforcement, such as withholding grant funds from noncompliant parties, has provided a noticeable increase in open access compliance and is a sure method of ensuring compliance.</td>
</tr>
<tr>
<td>Protectionism and Profitization</td>
<td>Protecting possible patents is important to universities and researchers.</td>
<td>Researchers’ attitudes about open access and open data vary from wanting to lock up the data to being “gung ho”, particularly when public funds are used. Philosophical ideals of sharing data and research for the greater good resonate more with younger faculty.</td>
</tr>
<tr>
<td>Concerns of stolen ideas and research</td>
<td>Concerns that others will steal ideas and research if data is shared before all research is conducted.</td>
<td>Researchers and funders agree on terms to embargo the data--data is stored in a managed data set and is made public after a certain time period (i.e., one year after the research is published).</td>
</tr>
<tr>
<td>Fears that private, identifiable information will be discovered</td>
<td>Human subjects research often call for anonymization of information. Also, for those situations where information can be drilled to an individual record that exposes personal information (i.e., poultry farm in Smith County when there is only one poultry farm in that county) is a concern.</td>
<td>Data can be cleansed and anonymize before it is uploaded. Careful attention to needs to be given to data sets where individual records exposes personal information.</td>
</tr>
<tr>
<td>ISSUE</td>
<td>THE CHALLENGE</td>
<td>COUNTER PERSPECTIVES OR RESOLUTIONS</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Technical expertise needed to create and implement data management plans</td>
<td>Not having the knowledge to create secure and effective data management plans is a limitation to advancement of open data. Funding agencies expect more secure plan for the data's longevity.</td>
<td>Librarians offer technical support and training. Some researchers don’t realize that storing research data on single hard drive is not effective way to secure and maintain data. Over time researchers will become more adept at data management.</td>
</tr>
<tr>
<td>Not understanding the universities’ licensing and ownership policies</td>
<td>Complicated contracts, technology transfer agreements, and policies make it difficult for faculty to understand and follow legal agreements. Some faculty don’t read the agreements about the rights to published articles and research data.</td>
<td>Various licensing options for re-use using Creative Commons, or other options, limit faculty from adopting open access and open data.</td>
</tr>
<tr>
<td>Moving toward cloud services may be problematic</td>
<td>Cloud services such as Amazon Web Services charges for withdrawal, not for depositing data.</td>
<td></td>
</tr>
<tr>
<td>Not having enough resources</td>
<td>Developing repositories, storing data and managing data sets, and providing secure infrastructure require resources that some universities do not currently have.</td>
<td>Using subject specific and 3rd party repositories relieves universities of some of this burden.</td>
</tr>
<tr>
<td>Unrealistic expectation that data will be saved for perpetuity</td>
<td>There are not many easy reliable and long terms solutions to save data forever.</td>
<td></td>
</tr>
<tr>
<td>Open data can lead to a problem with data quality</td>
<td>Researchers worry about the quality of the data and the risks to research when there are problems with data quality.</td>
<td>Openness gives opportunities to confirm or question research.</td>
</tr>
<tr>
<td>Some of the new open access journals lack quality</td>
<td>While not all open access journals lack quality, a few don’t adhere to quality peer review and standards which creates a dangerous precedent for research.</td>
<td>Research indicates citation rates for journals which have switched from a subscription based model to open access increased citation rates for top rated journals and decreased citation rates for bottom ranked journals, possibly increasing competition and emphasizing the importance of quality to citation rates.</td>
</tr>
</tbody>
</table>
GODAN conducted a study with the United Kingdom Department for International Development (DFID), United States Agency for International Development (USAID), and the Bill and Melinda Gates Foundation (BMGF) to review and compare open data policies of these organizations with Open Data Institute’s (ODI) checklist of policy elements. They interviewed and surveyed stakeholders and assessed the data quality of funded agriculture programs.

Both the GODAN donor study and the PUSH university assessment found that:
1. Grantees (which include researchers and implementing partners from universities) struggle to ensure data quality through effective data management and the provision of pathways for responsible data reuse.
2. Lack of proper budgeting and institutional open data policies hinder the process of opening up and publishing data correctly.
3. Lack of clear directives on where to publish and low awareness of how researchers can access the published data create barriers to compliance. This is especially true for researchers without a mature disciplinary repository. For example, pharmacy and veterinary medicine have strong options for sharing data within their field, but other disciplines do not have comparable choices.

The recommendations that GODAN offers in its study feed directly into the results to those found in this PUSH study, including:
1. Provide resources for researchers and program managers, including opportunities for peer-learning among donor organizations and research partners.
2. Incentivize researchers to publish data by rewarding quality data production and including data management support in long term budgets.
3. Foster a culture of change through openness, collaboration, and messaging at the highest level of authority.
4. Monitor policy implementation and data quality while improving process by designating these responsibilities.
5. Provide access to shared tools for data publication and consider platform selections.
Recommendations to Advance Open Data

Four organizations—PUSH, GODAN, AAU, and APLU—have produced reports and recommendations calling funders and universities to take action on open data policies and practices, urging universities, researchers, and funders alike to not fall behind in the open data ‘revolution’. PUSH’s recommendations reinforce the recommendations from these other organizations. PUSH calls upon universities to make changes that create stronger open data policies and infrastructure.

I. Communicate the benefits of open data.
Most librarians believed they could do a better job marketing the values of open data and data management, especially to aid in overcoming misunderstanding and doubts about open data.

Action:
Communicate the benefits of open access and open data.
Communications about the value of open data should include what it means to the public, research community, the university, and the individual researcher’s reputation. While open data can yield a public good, it can bring increased recognition to universities’ and faculty’s bodies of work. Open data also provides opportunities for collaborations and funding.

II. Create alignment between funders’ expectations and universities’ capabilities.
As the GODAN report indicates, funders are realizing that there are costs to sharing data1. Areas for alignment are:
- Include costs or partial costs of sharing and maintaining data sets in the project budget.
- Agree on a timetable for publishing findings and data sets, embargoing the data until articles are published.
- Delineate standards and protocols to link data and enhance its interoperability.

Actions:
Develop agreements between funders and universities that reflect solutions to open data challenges. In the request for proposals and budgeting, these entities can begin to agree on solutions as part of the research project process.

Conduct another GODAN open data summit. In the 2016 summit17, challenges were recognized and recommendations were made to move toward open data implementation in the agriculture and nutrition research community. The next summit should focus on creating synergies and alignment among researchers and funders.
Provide ongoing communications and conversations about open data opportunities and challenges. Through organizations like PUSH, GODAN, CGIAR, AAU, an APLU more sharing can be provided online and at conferences. There is need for universities and researchers to discover effective ways to develop open data policies, to build buy-in within the agriculture and nutrition research communities, and to observe working practices of effective open data.

III. Improve faculty’s ability to comply with open access and open data requirements.
While there are attitudes to support the greater good, it is clear that the movement of open access and open data within universities is being driven by funders’ requirements.

**Actions:**

Create a partnership between sponsored programs and libraries. Follow the model of a few universities which have developed a proactive policy and practice of coordination between the sponsored programs office and their libraries. Ensuring that there is communication between the office where contractual agreements are created and those who have technical expertise enables faculty and the university to comply with open data requirements.

Provide greater support for developing effective and smart data management plans. Universities, colleges, and libraries should encourage, facilitate, and support faculty in designing data management plans during the proposal stages. Support includes:
- providing information and recommending repositories and registries for agriculture and nutrition research; and
- building professional development capacity for researchers to learn how to design and implement open data.

Integrate an open data theory and practice course into master’s and doctoral curricula. Graduate students seem to be more aware of the need for open data and its potential. Providing education through a course that would include history of open access and open data, licensing options, data management design and planning, and repository and registry options establishes a foundation for future effective open data practices among new researchers.

IV. Refine and/or develop open access and open data policies.
Developing open data policies also means that there needs to be an assessment of current systems and culture. Policies should be supported by informed guidance, state-of-the-art tools, and expert technical support.

**Actions:**

Clarify licenses and ownership. Moving toward open access and open data policies requires clarification of ownership, transfer, and licensing. Clarification in policies should:
- address the ownership of faculty’s work and data collected in research;
- define when data are considered facts and when data are part of researchers’ work; and
- incorporate into policy statements how data can be shared by researchers.
Foster faculty buy-in for an open data policy. In developing open data policies, there must be acceptance from faculty. Interviewees were adamant about faculty buy-in being a necessary requisite to change and/or develop new policies.

Include adherence to open data strategies in tenure and promotion. The current structure for tenure and promotion does not encapsulate the potential benefits of open data for faculty nor for universities. An evaluation of open data incentives that can be included in the tenure and promotion process could be important going forward.

V. Create infrastructure.
Most of the universities involved in the study do not have enough resources in the library to handle the widespread use of open data. In fact, some interviewees said if there were an increase in open data requests, there would not be enough resources to support them. Likewise, over time, libraries may not have the infrastructure for long term maintenance and curation of the data sets.

Actions:
Build in more resources for data planning and management. With more and more funders requiring commitments to open data and mounting attitudinal pressure that public funded research is a public good, understanding open data and its management is becoming increasingly important. The Johns Hopkins structure provides an excellent model for developing cost systems of open data.

Use ODI open data resources. ODI provides several guides to accessing open data, anonymization and risk control of publishing open data, data licensing, identifiers for open data, and planning and budgeting for an open data initiative. Universities should consider these ODI resources:
- guides for developing an open data policy;
- guides to plan and budget for open data; and
- template to assess the maturity of the open data system.

In 2014 John Hopkins implemented a fee of $1,600 for faculty to upload small data sets (20GB) and 2% of the direct costs of the grant for large datasets (2TB) to recoup part of the cost of storing data and maintaining the infrastructure. The Library provides support and consultation throughout the project for 5 years and renewable after 5 years.

Fees are believed to be a reasonable charge for faculty, and yet these fees don’t cover the total costs of providing a repository.

Harmonize metadata, protocol, and standards. Data management specialists and librarians suggest that data be shared in nonproprietary formats and include documentation in metadata so that the data are discoverable, machine readable, and usable to others. As agriculture and nutrition research data registries and repositories mature, there is a greater need to use common language and harmonize the data across systems.
Appendix 1: Interview Questions

These questions were used as a guide for the interviews.

IDENTIFICATION SECTION
What is your name?
What is your university?
What is your department?
What is your title or position?
Do you know what PUSH is? What is your affiliation with PUSH?

CURRENT STATUS AND POLICY
Let’s start with policies and the status of open data at your university.
Does your university focus on open access or open data? Or both?
Where is your university currently with regard to open data?
Has open data been talked about or implemented university-wide or within departments?
Do you have a university-wide open data policy?
Do any departments have a policy? If so, which are they and can you supply us with the policies or give us names of someone who may have the policy?

REPOSITORIES AND PORTALS
Do you have a university-wide or departmental portal(s) or repository(ies) for sharing data?
Do you have any staff tasked with data management or uploading datasets to repositories?
Where have you registered your repositories?
What platforms are they on?
Why have you chosen these?
Which interoperability protocols and standards will these repositories support?
Which subject thesauri, controlled vocabularies, ontologies, or taxonomies --if any-- are used within your data repository/ies?

TRAINING AND SUPPORT
Is there anyone who is responsible for training and assisting faculty with uploading data and using data repositories?
Is there anyone who is responsible for deciding metadata for open datasets? Are metadata standards used? If so, which ones (refer to CGIAR as examples).
Do you have any datasets, portals, policies that could be useful to other universities and entities? Are they easy to find and access?

Please provide descriptions or links to these databases.
Do you know how researchers at your university feel about the policy (or lack of policy)?
What is preventing them from fulfilling the policy?
Do you feel there is good researcher awareness of Open Access/Open Data requirements at your university?

PROCESS
What was (or will be) your process for creating an open data policy or portal?
To whom would you give credit for initiating open data at your university?
Were there any tools or guides you referenced that were helpful?
How have the faculty and researchers reacted to the open data initiatives? Have most been onboard or is there resistance and concern?
Please describe how faculty have described the benefits and resistance of open data initiatives.
What are the incentives for faculty and researchers to follow the policy?
What are the challenges and barriers from a university perspective? From faculty perspective?

LESSONS LEARNED
Have there been any successes? What has been the most helpful thing in this process?
What have been the challenges? What barriers have you faced? How have you addressed them? Is there anything you wish you knew earlier in the process?

CURRENT NEEDS
Do you think it is important for faculty and researchers to share their research in an open data format? Why or why not?
Do you think faculty and researchers see data sharing as a top priority? Why or why not?
What resources or tools do you need or wish you had when creating this initiative?
Is there anything else you would like to share about this process or about your needs in this area?
Appendix 2: Terminology

DOI: digital object identifier is an “interoperable identifier” used on digital networks, allowing for unique identification to be given to research articles on the internet, making it easier to find articles (doi.org)

Metadata: data about other data. This data can provide various information including: title, abstract, author, date of publication, intellectual property rights, who has access, what is needed to preserve and archive a source, etc. (http://marciazeng.slis.kent.edu/metadatabasics/types.htm)

Open Access: usually the lowest “tier” of open data as you will see in the below definitions: “Open Access is the free, online availability of research articles coupled with the rights to use these articles fully in the digital environment.” Source: Scholarly Publishing and Academic Resources Coalition (SPARC)
  - Gold Open Access: open access publishing, is always free to the reader
  - Green Open Access: can archive pre-print and post-print (pre-peer review and post-peer review) or publishers version/PDF

Open Data: research data that is freely available on the internet permitting any user to download, copy, analyze, re-process, pass to software or use for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.

Repository
  - Data Repository: In this context a data repository is a destination, site, or place for a cyber collection of data through several entities (networks) or a singular entity. It is used to hold and separate data from a larger network or collection of data.
  - Institutional Repository: A repository for data, held and organized by an organization, business, university, or college.
  - Subject Specific Repository: A repository for data, organized by subject or field of study i.e. archaeology.

Team Science: a collaborative effort involving various researchers and scientists attempting to answer a research question. Often times involves collaborations across disciplines.
Citations


GODAN supports the proactive sharing of open data to make information about agriculture and nutrition available, accessible, and usable to deal with the urgent challenge of ensuring world food security. Organizations can join by registering at www.godan.info/become-a-godan-partner.

PUSH is a consortium of universities from around the world focused on inspiring collective action to end hunger and malnutrition, both locally and globally. University presidents and chancellors interested in joining should go to the PUSH website at pushtosign.org.

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI 10.17605/OSF.IO/E6T5F