
Plan Overview

A Data Management Plan created using DMPTool

Title: South Central NY Fruit & Vegetable Prescription Program (FVRx)

Creator: Baz Perry

Affiliation: Cornell University (cornell.edu)

Funder: Digital Curation Centre (dcc.ac.uk)

Funding opportunity number: 42585

Template: Digital Curation Centre

Last modified: 09-29-2020

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

South Central NY Fruit & Vegetable Prescription Program (FVRx)

Data Collection

What data will you collect or create?

Our data collection is divided into two categories: health care data protected under HIPAA regulations and participation and behavioral data used for program evaluation. Lists of data, along with their point of origin, responsible institution and use in evaluation are listed in Table D. The majority of data will be stored in spreadsheets. Databases, if used, will have clear export to CSV function.

With the exemption of some health data stored in electronic medical records, all the data listed is created by our Program Leadership teams and affiliates. Publicly available census data and regional and national public health data will be also used for analysis by our evaluation teams at Binghamton University and Cornell University. This data will be properly referenced when added to datasets.

How will the data be collected or created?

Data will be stored in a limited number of files that correspond to the following categories: Health Outcomes, Participation Rates, Program Feedback, Behavioral Surveys, Redemption Rates, Cost Measurements. During analysis, evaluators will then create new files with correlations between data points.

Because final datasets will all be sent to evaluators before storage, they will be responsible for organizing an aggregated data file for upload to the institutional repository. Improvements in standardization of data capture and data validation are expected through our Program Leadership collaboration between counties as well as with affiliated national organizations.

Documentation and Metadata

What documentation and metadata will accompany the data?

Data will include information about its origination, including but not limited to the following: Creator (the institution at which the data originated), Date Collected, and Methods Used (ex: phone interview, paper survey, online survey).

Metadata for datasets housed in eCommons at Cornell will include: Title, Author/Contributor, Place of Publication, Institution, Date Published, URL/DOI or identifier, Subjects, Keywords, Abstract/Summary, Language, Notes, Type (dataset).

Metadata for datasets housed in ORB@Binghamton will include: Title, Author/Contributor, Type (dataset), Date Published, URL/DOI or identifier, Keywords, Recommended citation.

Ethics and Legal Compliance

How will you manage any ethical issues?

Informed consent will be established by all health care providers working with participants. Participants will only be asked for permission to use their health data without identifiers.

Informed consent will also be established at CCE-Tompkins to inform participants about the behavioral and program research being gathered by staff members.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

All datasets will be available in the public domain.

Storage and Backup

How will the data be stored and backed up during the research?

Program teams have the opportunity to work with The Cornell Restricted Access Data Center (CRADC), which will help continue to update our data security plans and backup our research data sets, and provide access to a variety of data analysis software applications.

How will you manage access and security?

Privacy measures for participant data is a high priority and handled in accordance with HIPAA requirements. All identifiable health data is generated and recorded at health care providers and remains in their possession. Datasets are de-identified or aggregated before sharing with evaluation teams at Binghamton University and Cornell University using secure file transfer systems.

Behavioral and participation data that is not regulated under HIPAA will originate with and be recorded by Program Leadership teams and their affiliates (including behavioral survey data, redemption tracking data, web resources use data, evaluation questionnaires). This data may be shared between team members using the following security measures: keeping paper forms in a locked file cabinet at a secure institutional location that only team members can access, accessing digital datasets only on encrypted devices either within an institutional network or through a VPN (virtual private network), using secure file transfer systems to share datasets within the team and to share de-identified data and aggregated data with evaluation teams at Binghamton University and Cornell University.

Every individual collaborator with access to non-aggregated data will be listed and covered under their associated IRB. Only those listed as "researchers" on an IRB will have access to edit datasets for formatting and analysis.

Under the guidance of their IRB, evaluators will be responsible for using best practices in sharing this de-identified and aggregate data between team members and transferring datasets to their institutional repositories.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Aggregated data on both program administration and health outcomes is of lasting value to the ongoing work of the USDA and local and regional public health initiatives. This data may be used in the establishment of similar programs, for policy development, for revisions to future iterations of similar grants, and for evaluation of our program for continued support and collaboration. Therefore, all aggregated data from this program will be retained and shared publically and indefinitely.

What is the long-term preservation plan for the dataset?

All data generated through this program will be reported to Nutrition Incentive Program Training, Technical Assistance, Evaluation, and Information Centers (NTAE). Our hope is that NTAE will create a searchable database of prescription program research in the model of the FNEP database housed at <https://openpublishing.psu.edu/efnep/biblio>. In the interim and in addition to the work of NTAE, the dataset from Tompkins County will be housed in eCommons, Cornell's digital repository, found at <https://ecommons.cornell.edu/>. Datasets from Broome, Delaware, and Tioga counties will be housed in The Open Repository @ Binghamton, found at <https://orb.binghamton.edu/>.

Data Sharing**How will you share the data?**

Once datasets are formatted and analysis field are added, datasets will be made publically available through the following means: distribution through NTAE, public access on the corresponding institutional repositories, and links to datasets on Program Leadership websites.

Are any restrictions on data sharing required?

Once uploaded to institutional repositories and sent to NTAE, all datasets will be in the public domain and available without restrictions.

Responsibilities and Resources**Who will be responsible for data management?**

Datasets that are created within Program Leadership teams or affiliated organizations (ex: redemption sites, Cooperative Extension offices) are the responsibility of Program Leadership until they are shared with evaluators. Datasets that are created at health care providers are the responsibility of their own institutions until they are de-identified or aggregated and shared with evaluators.

Once evaluation teams take possession of datasets, they are responsible for data management at their institution (Binghamton University or Cornell University) and for ensuring that completed datasets are stored in their respective repositories with accompanying metadata. They are also responsible for notifying Program Leadership once datasets are uploaded into the repository so that they can be distributed to national and local websites. Program Leadership will work with evaluators to ensure timely completion of these tasks to meet reporting deadlines for evaluation teams.

What resources will you require to deliver your plan?

Program Leadership, health care providers and evaluators are equipped with the necessary software and storage capacity for this project. Both institutional repositories are provided free of charge for evaluators to store datasets for an unlimited duration and staff hours to complete data storage are being provided in-kind by evaluating institutions. Redemption sites in Broome, Delaware and Tioga Counties have been supplied with a redemption tracking system that requires ongoing subscription and technical support. The resources to continue this system (FMTracks) are reflected in the budget. Redemption sites in Tompkins County are served through the capacity of Cooperative Extension-Tompkins County and their redemption tracking system needs are also reflected in the budget.