Data Curation Profile – Food Technology and Processing / Food Preservation

Profile Author	Sonia Wade Lorenz & Lisa Zilinski				
Author's Institution	University of South Florida in Lakeland				
Contact	Lisa Zilinski, Izilinski@usf.edu				
Researcher(s) Interviewed	[name withheld], Professor of Food Science				
Researcher's Institution	University of South Florida in Lakeland				
Date of Creation	January 19, 2012				
Date of Last Update	August 23, 2012				
Version of the Tool	Data Curation Profile Toolkit is 1.0				
Version of the Content	Profile Content: 1.2				
Discipline / Sub- Discipline	Food Technology and Processing / Food Preservation				
Purpose	Data Curation Profiles are designed to capture requirements for specific data generated by a single scientist or scholar as articulated by the scientist him or herself. They are also intended to enable librarians and others to make informed decisions in working with data of this form, from this research area or sub-discipline.				
	comparison; however, they are designed to be flexible enough for use in any domain or discipline				
Context	A profile is based on the reported needs and preferences for these data. They are derived from several kinds of information, including interview and document data, disciplinary materials, and standards documentation.				
Sources of Information used for the profile	 An initial interview conducted on January 9, 2012. A worksheet completed by the scientist as a part of the interviews. A recording of the interview on January 9, 2012. A visit to the scientist's lab on January 11, 2012 				
Scope Notes	The scope of individual profiles will vary, based on the authors' and participating researcher's background, experiences, and knowledge, as well as the materials available for analysis.				
Editorial Notes	Any modifications of this document will be subject to version control, and annotations require a minimum of creator name, data, and identification of related source documents.				
Author's Notes The Food Science and Nutrition data curation profile is based on the an interview completed with the scientist working in this research are discipline. Some sub-sections of the profile may be left blank; this or there was no relevant data in the interview or available documents us construct this profile.					
Permissions	issions The researcher has reviewed the completed data profile and confirmed the accuracy of the information. The researcher gives consent to allow the profile be published on Purdue's website for completed data profiles.				
URL	http://datacurationprofiles.org				
Licensing	This work is licensed under a Creative Commons Attribution Image: Common State S				

Brief summary of data curation needs

The scientist is conducting a series of experiments on food preservation for the Department of Defense. While the data is valuable, it cannot be publicly shared without permission; privacy and security are high priorities.

The data is currently stored in multiple locations including on the lab computers, the scientist's home, and an external hard drive. The scientist also maintains a strict schedule for backing up the material. The scientist is interested in the possibility of being able to securely store data in a digital asset management system that could be accessed remotely; however, security and privacy are her main priorities.

Overview of the research

Research area focus

The scientist describes this project as having one global goal with three objectives. Globally, the team is creating a tool to allow the Department of Defense to ship MREs (Meals Ready to Eat) with the expiration information included. MREs undergo extreme changes in climate therefore an expiration tool is required.

Objective 1:	Create a handheld RFID (Radio Frequency Identification Device) that will
	measure the time and temperature profile of the MRE.
Objective 2:	Evaluate quality/shelf life of MREs in adverse conditions by:
-	A.) evaluating color, composition and texture of food and
	B.) evaluating taste of food.
Objective 3:	Evaluate quality/shelf life of fresh fruits and vegetables delivered to
	troops.

The scientist participating in the data profile is focusing on Objective 2 part A exclusively. For Objective 2 part A, the scientist will subject seven different types of MREs to five different environments: 44°, 84°, 100°, 120°, and 140° Fahrenheit. Samples of the MREs are tested incrementally throughout the experiment. The tests measure color, texture and composition. The composition testing measures the following: ph levels, acidity, soluble solids content, sugars, ascorbic acid and fat oxidation. With the information from all three objectives, the scientists will be able to create a tool to determine the quality and remaining shelf life of food delivered to soldiers.

Intended audiences

Other researchers working with MREs may be interested in the data. The findings could be applied to commercial food products. Civilian food manufacturers, packing companies, distributors and retailers may be interested in the findings. However, per the scientist's agreement with the Department of Defense, the data may not be made publically available.

Funding sources

The research is primarily funded by the Department of Defense. The University of South Florida in Lakeland also contributes to the project by providing the facility within which the research is conducted.

Data kinds and stages

Data narrative

In order to create a tool that will allow MREs to be shipped with the expiration information included, expiration data must first be generated and collected (Objective 2A). The scientist will subject seven different types of MREs to five different environments: 44°, 84°, 100°, 120°, and 140° Fahrenheit. Samples of the MREs will be tested incrementally throughout the experiment until the MRE is no longer fit for consumption. The tests will measure color, texture and composition. The composition testing will measure the following: ph levels, acidity, soluble solids content, sugars, ascorbic acid and fat oxidation. To obtain this data, the scientist utilizes several machines: a Texture Analyzer, the Minolta Chroma Meter, High Performance Liquid Chromatogrphy (HLPC), Titrator, PH Meter and a Refractometer.

The results are automatically exported from the above listed machines into Excel files. Once testing ends, the scientist will review and analyze the data. In August of 2012, the scientist plans to import the Excel files into SAS for the statistical analysis. In December of 2012, the scientist plans to use Sigma Plot to generate the necessary graphs and tables for her research. In February of 2013, the scientist expects that the project will be complete and she will be able to write the finalized reports and articles to summarize her findings.

The scientist has a rigorous method for saving and backing up her files to be certain nothing is lost. Each night the files are saved on the computer's hard drive and an external hard drive. At the end of each month the files are backed up on a separate hard drive and on a different computer housed at a separate location.

Data		# of Files /				
Stage	Output	Typical Size	Format	Other / Notes		
Primary Data						
Raw Data	Photos and data generated from Texture Analyzer and Minolta Chroma Meter	MB/JPG	Exponent, Utility 4 software. Pictures	Data is available on analysis machines and photos are taken to record the food's appearance.		
Export	Excel Files	700 files/ 40KB	Excel	Data is automatically exported from the machines to Excel.		
Statistical Analysis	Data is organized by feature (temperature, replicates, etc)	Unknown file size	SAS	The scientist will compile Excel data and execute statistical analysis using SAS		
Graphs / Tables Generated	Graphs and Tables	Unknown: 500+ graphs	Sigma Plot	Analyzed data is imported from excel into sigma plot to create graphs		
Finalized	Graphs and tables with an article.	Unknown file size	Graphs, Pictures, Tables & MS Word	The final report will include previously generated graphs and tables as well as a written document/article.		

The data table

Target data for sharing

The scientist is able to publish articles about her research but the data cannot be publicly shared without permission due to the privacy restrictions required by her funding agencies.

Value of the data

The scientist advises that the data could hold value for other researchers doing work on MREs and that her findings could be applicable to commercial foods. The concepts learned in this research could be applied to civil food retailers, distributers, and preservation.

Contextual narrative

The scientist emphasized that even though others could benefit from the data, the regulations required by the Department of Defense prevent the data from being made public without permission.

Intellectual property context and information

Data owner(s)

The Department of Defense is the sole owner of the data according to the scientist.

Stakeholders

The University of South Florida in Lakeland provides the facility within which the scientist conducts her research. The scientist also advised that she has research assistance working with her on this project. However, the Department of Defense has sole ownership of this project as this is considered in the same manner as a work for hire.

Terms of use

The scientist states that her immediate collaborators may have access to the data but the data cannot be made publicly available without permission.

Attribution

Because the data cannot be publicly shared without permission, attribution is not relevant.

Organization and description of data (incl. metadata)

Overview of data organization and description (metadata)

First, the data is populated by each machine and is automatically organized in Excel sheets. These excel sheet describe the features of the food products including date, temperature, color, ph levels, acidity, soluble solids content, sugars, ascorbic acid and fat oxidation. Toward the end of the project, the scientist will examine the Excel data using SAS and import the analyzed data into Sigma Plot to create graphs. Metadata is not currently used with this dataset.

This data set is well organized and well protected, but the scientist did express interest in being able to access and search the data from offsite locations. Assuming all required security measures could be met, the scientist advised that it would be very helpful if both she and her team could access the data sets remotely. Metadata is not currently used with this dataset. However, if the data was stored in a database that could be accessed remotely, metadata could be employed to make the dataset easier to search.

Formal standards used

No formal metadata standards, ontologies, or controlled vocabularies have been applied to this data. The scientist organizes her data by date.

Locally developed standards

The scientist organizes the Excel sheets in chronological order and by MRE sample groups. With this structure, she can see how each MRE sample is impacted over time.

Crosswalks

N/A

Documentation of data organization/description

The scientist did not mention any existing papers or documents that detailed standard operating procedures. The excel spreadsheets detail all of her data automatically.

Ingest / Transfer

The scientist indicated that because the bulk of her data is already in common formats (Excel), she did not anticipate needing any unique functionality from the repository. The scientist expects that the repository would automatically be able to read basic excel and word formats.

However, the ability to personally submit the data in single and batch uploads to the repository is a high priority for the scientist. Because of privacy concerns, the scientist requires complete control over the data including when and how the data is uploaded in the repository.

Sharing & Access

Willingness / Motivations to share

The scientist indicates that this project was commissioned by the Department of Defense and as a result, the dataset cannot be made public without permission.

Embargo

N/A

Access control

The scientist requires complete control over the data to ensure that she maintains the privacy guidelines set forth by the Department of Defense. She is interested in tiered permissions of authority because that would enable her team to have remote access to the data. The scientist also believes that tiered permissions would enable her team to work more efficiently. However, she must maintain primary control and access to the data cannot be extended outside her research team without permission.

Secondary (Mirror) site

The scientist indicates that a secondary/mirror site at a separate geographic location is a medium priority.

Discovery

Discovery is not important to the scientist because the data cannot be shared without permission.

Tools

To obtain this data, the scientist utilizes several machines: a Texture Analyzer, the Minolta Chroma Meter, High Performance Liquid Chromatogrphy (HLPC), Titrator, PH Meter and a Refractometer. The results are automatically exported from the above listed machines into Excel files.

While no additional tools would be necessary to be able to read or use the excel files, a data profile would aid significantly in being able to interpret and give context to the files.

Linking / Interoperability

Linking/Interoperability is not important to the scientist.

Measuring Impact

Usage statistics & other identified metrics

Usage statistics & other identified metrics are not important to the scientist because the data cannot be shared without permission.

Gathering information about users

N/A

Data Management

Security / Back-ups

Currently, the data is kept on the lab's computer hard drive and an external hard drive (these are updated manually with each new dataset collection). At the end of each week, the lab computer and external hard drive are backed up by the main lab computer. At the end of each month, the main lab computer is backed up to final external hard drive kept at a secondary location.

While the scientist is confident in this method, she is interested in being able to use a data repository as an offsite means for backing up her data.

Secondary storage sites

A Secondary storage site and a secondary storage site at a different geographic location are medium priorities for the scientist.

Version control

Version control is a high priority to the scientist.

Preservation

Duration of preservation

The scientist indicated that her data should be stored for 5 years.

Data provenance

Documentation of any and all changes made to the data over time is a medium priority for the scientist.

Data audits

The ability to audit the dataset is a medium priority for the scientist.

Format migration

The ability to migrate the dataset into new formats over time is not a priority for the scientist.

Personnel

Primary data contact

[name withheld]

Data stewards

Lisa Zilinski Business Librarian University of South Florida in Lakeland

Other contacts

N/A

Sonia Wade Lorenz Library Assistant University of South Florida in Lakeland