

Original Article

USDA's Nutrient Databank System – A tool for handling data from diverse sources

D.B. Haytowitz*, L.E. Lemar, P.R. Pehrsson

Nutrient Data Laboratory (NDL), Beltsville Human Nutrition Research Center, USDA-ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA

ARTICLE INFO

Article history:

Received 10 January 2008

Received in revised form 24 June 2008

Accepted 13 January 2009

Keywords:

Food composition

Database

Nutrient data

Food item table

Nutrient value table

Data compilation

USDA Nutrient Databank System

NDBS

Nutrient Data Laboratory

NDL

Dr. Atwater

What We Eat in America

NHANES

ABSTRACT

Key features of USDA's Nutrient Databank System (NDBS) allow processing of food composition data from diverse sources, including USDA's National Food and Nutrient Analysis Program, the food industry, scientific literature, and food labels. The Nutrient Data Laboratory (NDL) designed the NDBS as a three-tiered ("Initial", "Aggregation", and "Compiled") data management system to facilitate handling of data. Raw data and documentation (data source, sample description, sample handling, and analytical methods) are migrated into the Initial module. NDL scientists compare new data with old values and decide how to combine the initial data into aggregated data. In the Aggregation module, data can be grouped and weighted by parameters such as study, source, and market share. Depending on the type of data, various statistical algorithms are used to generate statistics, such as mean, standard error, number of data points, and error bounds. In the Compiled module, food names are finalized and common measures selected. Nutrient profiles are developed and missing nutrients/food components are imputed according to standardized scientific principles. A formulation application employing linear programming techniques, estimate, formulations for commercial foods and nutrient profiles based on the nutrient content of ingredients and target values derived from label information. A recipe application calculates nutrient profiles based on ingredients and their known proportions, allowing for the application of food yield and nutrient retention factors. The NDBS automatically documents how each value was derived and incorporates quality control checks at all levels. Prior to release, the completed nutrient profiles are reviewed by NDL scientists and, if approved, disseminated. The NDBS brings together a number of stand-alone modules and applications into one integrated system allowing the management of ~7500 food items for up to 140 nutrients/food components. Data points and documentation are managed and maintained in one place, providing an "audit trail" for each data point. The NDBS contains algorithms to assign confidence codes using NDL's data quality evaluation system. The NDBS permits the annual release of reliable data for a comprehensive set of nutrients/food components for a wide variety of foods on NDL's Web site: <http://www.ars.usda.gov/nutrientdata>. Through these releases, NDL provides food composition data for researchers, diet and health professionals, and consumers, including the "What We Eat in America" component of the National Health and Nutrition Examination Survey (NHANES).

Published by Elsevier Inc.

1. Introduction and historical background

The U.S. Department of Agriculture (USDA) has maintained tables of food composition for over 115 years, since the pioneering work of Atwater (Atwater and Woods, 1892). Published in 1892, the Atwater table of 178 food items contained data on five proximate components (water, protein, fat, total carbohydrates and ash), kilocalories (called fuel in the Atwater table), and refuse. Dr. Atwater's data sheets (Fig. 1) are currently maintained in the Special Collections Section of the National Agricultural Library in Beltsville, MD (<http://www.nal.usda.gov/speccoll>).

The historic tables are similar in content to the electronic spreadsheets used today for many aspects of the work. This work is currently conducted in Beltsville, MD by the Nutrient Data Laboratory (NDL), which is part of the Beltsville Human Nutrition Research Center of the Agricultural Research Service of the USDA.

With the expansion of the number of nutrients/food components included in the tables, the increased quantity of data, and the need to capture more information about the food samples, it became clear that computerizing the data would facilitate the process of compiling tables of food composition. The first Nutrient Databank System (NDBS) ran on the USDA's mainframe computer and was written in the COBOL programming language (Table 1).

In the 1980s, the system was redesigned and written in a new programming language, PL1. Over the next decade, the NDL developed a number of stand-alone applications to allow

* Corresponding author. Tel.: +1 301 504 0714; fax: +1 301 504 0713.

E-mail address: david.haytowitz@ars.usda.gov (D.B. Haytowitz).

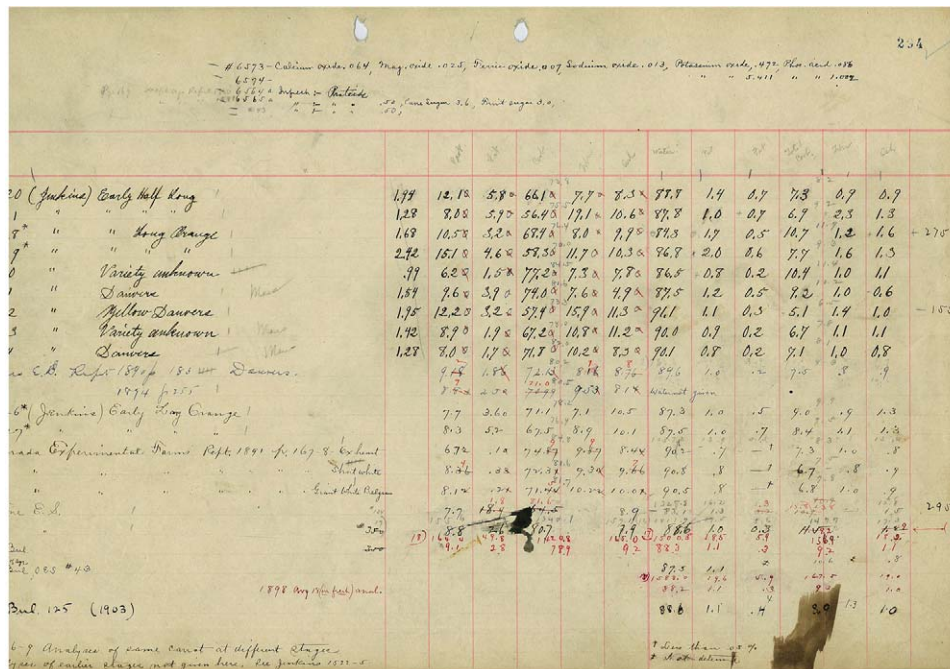


Fig. 1. Sample of Dr. Atwater's worksheets.

expansion of its database and to accommodate new data needs. For example, shortly after the completion of the 1980s databank system, a formulation estimation application was developed using the General Algebraic Modeling System to perform linear programming calculations (Marcoe and Haytowitz, 1993). This stand-alone application enabled nutrient estimates to be calculated by recipe using estimated ingredient percentages when analytical data were not available. This application was invaluable in allowing NDL to expand data for emerging nutrients of public health significance to be used in food intake surveys and epidemiological work.

By the mid-1990s, the need for a totally redesigned system was recognized. This report describes features of the latest version of the USDA's NDBS, the Architecture and Integration Management: Nutrient Databank System (AIM_NDBS). Though this system has existed in two architectures, one Client-Server (Oracle Client) and the other Web-based (Oracle Web), the features and functionality are the same.

2. System overview

The goal of the AIM_NDBS development was to provide an integrated system to handle acquisition, evaluation, compilation, storage, and dissemination of food composition data. The NDBS was developed to integrate all critical staff functions associated with nutrient data work, particularly those functions associated with development and dissemination of the USDA National

Table 1
Timeline and computer applications used in developing the USDA Nutrient Databank System.

Year	System	Platform and software
1976	NDBS	Mainframe – COBOL
1980	NDBS	Mainframe – PL1
1990	Master Database	Corel® Paradox/® – Novell® Server
1997	AIM_NDBS	Windows NT Server – Oracle® Client-Server
2005	AIM_NDBS	Windows NT – Oracle® Web-based application

Nutrient Database for Standard Reference (SR) (NDL, 2007a). Planning for the system began in 1997.

The project to develop the new system was divided into three major phases:

- Phase 1: Determination of system requirements.
- Phase 2: Hardware and software selection.
- Phase 3: Final system design, development, procurement, testing, and installation.

The AIM_NDBS utilizes a relational database comprised of 271 tables containing various sets of data to support the system, residing in an Oracle® database.

The NDBS was designed as a three-tiered data management system, comprising three modules:

1. The Initial Food Item module – Detailed nutrient/food component, weight, and physical component (i.e., part of plant or animal determined by dissection, including flesh, peels, bones, etc.) values are entered and food item description and methodology information are documented.
2. The Aggregation Food Item module – Individual nutrient/food component, weight, and physical component data for similar food items can be aggregated.
3. The Compiled Food Item module – Missing nutrient/food component values are imputed using standardized procedures, including recipes and/or formulations, and the food item profile is finalized for dissemination.

In addition, the NDBS has a variety of support areas which can be categorized as:

1. Utilities – Applications that assist NDL in maintaining and updating a variety of support tables; provide factors necessary for many unit conversions; and provide functionality for rating data quality, calculating nutrient retention factors and estimated cooking yields, and disseminating several databases, including SR.

Download English Version:

<https://daneshyari.com/en/article/1218953>

Download Persian Version:

<https://daneshyari.com/article/1218953>

[Daneshyari.com](https://daneshyari.com)