



# Data Warehouse and Environmental Public Health Tracking Business Plan December 2011

Final Report for the IDPH Data Needs Assessment



---

Iowa Department of Public Health— 321 E12th St, Lucas State Office Building, Des Moines, IA 50319



---

## Project Staff and Contact Information

---

Meg Harris, MPH, MPA  
Data Warehouse Coordinator, Project Manager  
Bureau of Communication and Planning  
[Meghan.harris@idph.iowa.gov](mailto:Meghan.harris@idph.iowa.gov)  
515-725-2179

Michelle Lenerz  
Data Warehouse Technical Lead  
[Michelle.lenerz@idph.iowa.gov](mailto:Michelle.lenerz@idph.iowa.gov)  
515-725-2203

Robert Walker, MPH  
Environmental Health Specialist  
Environmental Public Health Tracking Program  
Bureau of Lead Poisoning and Prevention  
[Robert.walker@idph.iowa.gov](mailto:Robert.walker@idph.iowa.gov)  
515-281-0908

Tim Wickam, MPH  
Community Health Consultant  
Environmental Public Health Tracking Program  
Bureau of Environmental Health Services  
[Timothy.wickam@idph.iowa.gov](mailto:Timothy.wickam@idph.iowa.gov)  
515-281-7462

Caitlin Lueck  
Data Warehouse Intern  
Bureau of Communication and Planning  
[Caitlin.lueck@idph.iowa.gov](mailto:Caitlin.lueck@idph.iowa.gov)

---



---

## CONTENTS

---

Background and overview of business needs .....	5
Data use and accessibility .....	8
Priority datasets for inclusion in 2011-2012 iteration of the Data Warehouse .....	8
Confidentiality requirements and data suppression.....	12
Display and visualization .....	13
User roles.....	16
Data processing, ownership policies, data acquisition and migration timing .....	19
Reporting .....	21
Metadata standards.....	23
Financial sustainability.....	25

### Tables and Figures

Table 1. IDPH staff self-reported challenges with access to data .....	7
Table 2. Percentage of IDPH programs with confidentiality policies, and those reviewed by the Iowa Attorney General ....	12
Table 3. Percent of analytic competencies used by IDPH programs.....	14
Table 4. Stakeholder rankings of existing functions of the DWH .....	15
Table 5. Stakeholder rating of the ease of the overall use of the DWH .....	15
Table 6. Preferred options for advanced users, research and patient-level data access as reported by stakeholders .....	17
Table 7. Annual load schedule for the DWH.....	20
Table 8. Use of data in monitoring population health, reported by average users .....	21
Table 9. Use of non-statistical features as reported by stakeholders .....	22
Table 10. Type of DWH use as reported by stakeholders .....	22

---

Figure 1. Percentage of types of data reported by IDPH staff, stakeholders and average users .....	9
Figure 2. Percentage of data users need access to as reported by IDPH staff, stakeholders and average users.....	11
Figure 3. Percentage of geographic display level preferred, as reported by average DWH users.....	13

---



Figure 4. Average users indicating whether there is a need to bring together multiple datasets in the DWH ..... 14

Figure 5. Distribution of frequently used analytic programs as reported by average DWH users ..... 15

Figure 6. Employer categories reported by stakeholders ..... 17

Figure 7. Percentage of formats for release of data as reported by IDPH programs, stakeholders and average users ..... 21

**Appendices**

Appendix A- EPHT grantee portal standards and recommendations, April 2009 ..... 13

Appendix B- Grantee portal descriptions by state, 2011 ..... 13

Appendix C- Metadata content guidance document ..... 24



## Background and Overview of Business Needs

The Iowa Department of Public Health (IDPH) Data Warehouse (DWH) project was initiated in 2008 with the initial intent of providing a data resource for external public health partners. Local public health agencies (LPHA) were targeted as the primary stakeholders when the application was rolled out in 2010. At that time, LPHA were working to complete Community Health Needs Assessments and Health Improvement Plans (CHNA/HIP), a planning activity formally conducted every five years. LPHA had a need for access to county-level data for a myriad of health indicators for planning as well as evaluating progress on indicators over time. In a limited capacity, agency staff worked with IDPH project leads to determine the content of the warehouse based on these health indicators. LPHA was integral in designing the county health snapshots currently available in the warehouse.

The DWH is a repository of point-in-time public health events and does not include individual identifiers (i.e. SSN, Medicaid) that may allow the tracking of an individual over time. Data are available at the county-level. Data are refreshed annually following a review and close-out process unique to each dataset. These data are final, non-provisional datasets and include the following subject areas: birth events, death events, inpatient hospital discharge events, census population estimates, and Behavioral Risk Factor Surveillance surveys from 2000 through present.

IDPH staff and public health partners may request access to the DWH at any time through the DWH Coordinator. All users are required to sign a confidentiality agreement as confidential and identifiable information can be derived from the DWH.

Users are categorized into one of two roles: basic or advanced users. Basic users gain access to the DWH data by logging into a secure SharePoint site where they can view and download content, but are not allowed to contribute content. This access requires Active Directory (AD) authentication as data are not suppressed and could potentially be used to identify an individual or group of individuals. Once in the basic user site, users can choose to view County Health Snapshot PDF reports or corresponding text data files for a select county by year or view a predetermined set of reports by subject area. These subject area specific reports are pivot tables contained in Excel workbooks which access the DWH cubes via Excel Web Services. Each report has a predetermined set of available data attributes as determined by the Excel workbook publisher. Copies of workbooks can be downloaded locally; however, data connections are severed, providing a point-in-time snapshot of the data. Each workbook contains a worksheet for indicator information, county-level data and state-level data. Measures are reported as counts, rates and percentages.

Advanced users are granted additional unrestricted access to browse all subject area cubes using an SQL Server Analysis Services (SSAS) connection in Excel. In order to access this connection, an advanced user must be locally connected to the IDPH network or remotely connect to the IDPH network using a VPN and token-authentication. Excel connection templates are available to advanced users on a mapped network drive (typically W:\). Whereas basic users are limited to a predetermined set of data attributes, advanced users have access to all attributes in each cube and browse the cubes using Excel pivot tables. Advanced users can then utilize all the native Excel functionality for charting and graphing. Additionally, using pivot table functions, users can change the display of measures. Data are not suppressed for advanced users.

Currently, the IDPH Information Management team is evaluating efforts to rebuild the basic user site on the SharePoint 2010 platform and recreate the set of user reports using Reporting Services in place of Excel Web Services.

---



## Overview of business needs

In 2010, Iowa became the 22nd state in the CDC's Environmental Public Health Tracking Network (EPHT). The utilization of data and information regarding health outcomes, environmental hazards, and human exposure/biomonitoring, or a combination of them, provide important information for public health practice and comprise environmental public health tracking activities. The availability of these types of data in a standardized tracking network will enable researchers, public health authorities, healthcare practitioners, and the public to begin to understand the possible associations between the environment and adverse health effects. A key characteristic of environmental public health tracking is the emphasis on data integration across health, human exposure, and hazard information systems. The outcomes of this data integration reach beyond Iowa and to 22 other states working on the tracking initiative. Since the national program is structured to require participating states to display data in the form of set indicators, data for those states are comparable and familiar to users.

The EPHT Program has a mandate to provide data to both the general public and expert users. These two types of users have different levels of understanding on health and environmental issues. Because of this difference, IDPH needs to provide both public and secure portals. The EPHT Program defines a portal as a web site that functions as the interface for accessing information on the EPHT Network. Portals at the national and state levels will allow for public access to unrestricted data and role-based/use-based access to restricted data. The public portal will make appropriate data available to a general user. While public portals may have static web pages and reports, these portals are expected to offer additional features such as ad hoc queries. The secure portal makes more detailed information available only to authorized users. Authorized users could include state and local health department staff, staff of other government agencies, academic researchers, or non-profit advocacy groups.

State portals are an important part of the EPHT Network. While the fact that state portals will have to meet the state's unique style and format guidelines is understood, certain other functionality will be required as part of the overall EPHT Network. For example, each state portal needs to provide on the state's main page a hyperlink to the National Portal. That state's organizations may require a particular header and style is also understood. Nevertheless, the hyperlink to the National Portal should be in a manner consistent with the state's style guidelines.

Each portal should employ a consistent, structured flow and should maintain visual consistency. The flow and visual make-up of a portal may be dictated by the state, but should still be applied consistently throughout the portal. In addition, page contents should be organized to promote easy identification of information.

The EPHT project involves the integration of environment and health data into the DWH, configuring Iowa data in the format of 78 nationally derived indicators, and the eventual public access to these indicators<sup>1</sup>. The indicators must be made accessible to the public through a public access and secure portal prior to the summer of 2013.

In 2011, the DWH and EPHT projects were integrated so that the expansion of warehouse functionality will include all EPHT objectives. Strategic planning for both projects resulted in several well-defined goals for 2011 and 2012.

- *Data needs assessment*
  - In order to assess the data needs of existing and future users of the DWH, as well as prioritize the objectives of EPHT, IDPH conducted a data needs assessment. The assessment was significant effort; IDPH staff,

---

<sup>1</sup> <http://www.cdc.gov/nceh/tracking/>

---



external users and existing users were surveyed. The results of the assessment provide quantitative and qualitative support for the content of this business plan.

- *Business requirements for DWH modifications and construction of a public portal*
  - IDPH will be working with the Department of Administrative Services and the Information Technology Enterprise (DAS/ITE) to draft business requirements based on this plan.
- *Construction of the public portal and modifications detailed in the business requirements*
  - Details of this objective will be part of writing the business requirements.

### Plan structure, content and assumptions

Challenges with data access and management have impacted efficiency and effectiveness of IDPH program staff. In the data needs assessment, more than 30 percent of staff indicated project or reports delays due to issues in data access (Table 1). Other tasks that were reportedly impeded by data access problems included customer service, program evaluation and the ability to comply with a data request.

**Table 1. IDPH staff self-reported challenges with access to data**

How have challenges with data access impacted your program?	Count	Percent
Delayed projects or reports	27	32%
Inability to comply with a request	21	25%
Inability to evaluate programs or performance	19	22%
Delays or impacts on customer service	18	21%
Delayed grant applications	3	4%

The results of the assessment used in this plan are displayed in three ways- 1) IDPH user, 2) stakeholder (includes existing users and external user group), and 3) average user which is the average need of both stakeholders and IDPH staff. The percent needs were placed into priority categories as follows:

- High priority- 70-100 percent of users indicate need; or priority of the EPHT project
- Medium priority- 30-69 percent of users indicate need
- Low priority- 0-29 percent of users indicate need

The DWH application should provide a way to streamline data management and access, leading to more efficient use of staff time and continuity of data resources throughout the department. External users will benefit from the ability to view and use data that was previously inaccessible, and without the need for making a request of IDPH staff. The overarching goal of this project is to make the DWH a more useful, sustainable tool for all existing and future users while meeting the requirements of the EPHT grant.



## Data Use and Accessibility

### Priority datasets for inclusion in 2011-2012 iteration of the DWH

#### High level priorities

Prioritization for inclusion of new datasets is primarily driven by the needs of the EPHT project. Current standards of the DWH dictate that years of available data must include 2000 through the most recent year available. Cancer and Birth defects have longer lag times due to retrospective case finding. When available, 2010 will also be migrated. Future years will be moved into the DWH according to a yet-to-be determined annual load schedule.

The state portal should provide access to environmental, exposure, and health effect data. This includes the Nationally Consistent Data and Measures<sup>2</sup> (NCDMs) which represent EPHT standards for core measures for asthma, myocardial infarction hospitalization, PM 2.5, ozone, water, vital statistics, childhood blood lead, cancer, birth defects, and carbon monoxide poisoning. These standards were developed by EPHT working groups, or Content Workgroups, and adopted by the CDC national EPHT program.

For the purposes of EPHT the following datasets are the source of required indicators and measures and are not already part of the DWH:

Air Quality – The data for this content area include monitored and modeled data. The data are collected by the Iowa Department of Natural Resources (IDNR); the data are published by IDNR via the Environmental Protection Agency (EPA) and are accessible through an FTP site. IDPH has the raw data files for monitored data. The modeled dataset is produced by the National Tracking Program at CDC in collaboration with the EPA, and is provided by the National Tracking Program for use in participating states. Data are displayed in daily and annual average time periods. Data is consolidated for display at the county level. The EPHT program will work with IDNR in determining who will assume the role of data owner/steward in future years.

Community Water Supply - The data for this content area are collected by the IDNR. Data includes descriptive information for all Community Water Systems in Iowa, and water quality test results for multiple contaminants. Data will be provided to IDPH by IDNR under a data sharing agreement. These data are captured at the Water System Level; the frequency of collection is typically quarterly. The water dataset will include identifying information of the water supply system. The CDC Content Workgroup is currently working on best practices for how to best display the water data on maps. The current DWH standard is county-level display. The EPHT program will work with IDNR in determining who will assume the role of data owner/steward in future years.

Birth Defects – The data for this content area are collected by the Iowa Registry of Congenital and Inherited Disorders (IRCID) at the University of Iowa. Data included in the DWH will be aggregated, meaning it will not contain case level data or personally identifiable information. The aggregated data will still require the ability to suppress cells for small numbers. Data will be provided to IDPH by IRCID under a data sharing agreement. The EPHT program will work with the University of Iowa in determining who will assume the role of data owner/steward.

Cancer – The data for this content area are collected by the Iowa Cancer Registry (ICR) at the University of Iowa. Data

---

<sup>2</sup> Specific information on the indicators and measure can be found in "Centers for Disease Control and Prevention Recommendations for Nationally Consistent Data and Measures within the Environmental Public Health Tracking Network"

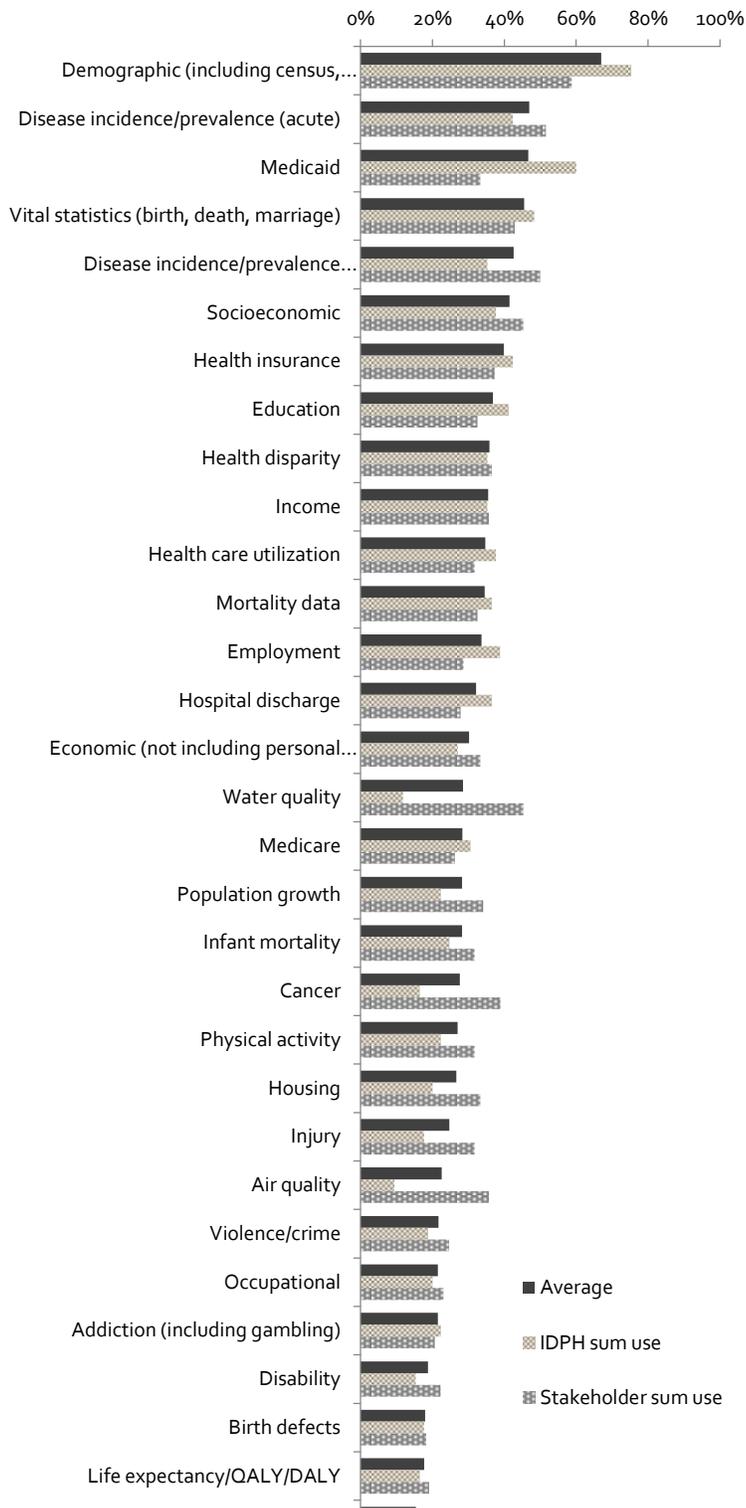


Figure 1. Percentage of types of data reported by IDPH staff, stakeholders and average users

included in the DWH will be aggregated, meaning it will not contain case level data or personally identifiable information. Displaying the aggregated data will still require the ability to suppress cells for small numbers. Data will be provided to IDPH by ICR under a data sharing agreement. The EPHT program will work with the University of Iowa in determining who will assume the role of data owner/steward.

Child Blood Lead - The data for this content area are collected and managed by the IDPH Bureau of Lead Poisoning Prevention (LPP). Data included in the DWH may be comprised of case level data, but will only display blood lead level readings and will not contain any personally identifiable information. The bureau chief for the LPP will serve as the data owner, and an Environmental Health Specialist working with the EPHT program will serve as the data steward.

*File format*

As indicated by the assessment, 88 percent of users utilize Excel, and csv file types would satisfy the needs of programs using other analytic programs.

**Medium-level priorities**

The needs assessment identified datasets users frequently utilize and need. In an effort to centralize common data files, the most frequently used files should be hosted in the DWH. An average of 40 percent or more of all users have a medium-level need to access the following datasets (Figure 1):

- Census
- Disease incidence/prevalence (acute)
- Medicaid
- Vital statistics (birth, death, and marriage)



- Disease incidence/prevalence (chronic, excluding cancer)
- Socioeconomic
- Health insurance

Several of these datasets are already in the DWH.

The most commonly used data are basic **census**. County-level display is a DWH existing standard, and all rates calculated in the warehouse use census estimates. Basic county and state level census files should be included in the warehouse. Parameters for these files are:

- Census estimate and decennial census files from 2000-2010
- By county, race and ethnicity
- By single year age when available
- In both excel and csv file format

Provision of census data is a simple task that may be completed by IDPH DWH project staff. Updates to census data should be made in coordination with the annual data load.

**Acute disease data** are used by an average of 47 percent of DWH users (Figure 1). Though only an average of 10 percent of users indicate a need to access the data, providing aggregated event counts by county for acute disease data would grant access to those in need as well as centralize access for those using the data often (Figure 2). Internal discussions with the Center for Acute Disease Epidemiology should begin before the requirements writing phase of this project.

**Medicaid data** should be considered for inclusion in this iteration of warehouse modifications. IDPH and Department of Administrative Services, Information Technology Enterprise (DAS/ITE) should begin discussions with Department of Human Services (DHS) for sharing Medicaid data. Negotiation of data sharing with DHS may need to be routed through the IDPH Ethics Committee and the Attorney General's office. IDPH needs to explore how DWH users utilize Medicaid data. Some programs require patient-level Medicaid files for matching to other IDPH datasets. Other programs and stakeholders may only need aggregate counts. Confidentiality standards must be maintained with the importation of this dataset, and should be determined when the data sharing agreement is drafted.

**Vital statistics data**, excluding marriage and divorce events, are already in the DWH.

**Disease incidence/prevalence for chronic conditions**, excluding cancer, is used by an average of 43 percent of DWH users (Figure 1). These data are available in the statewide inpatient discharge (SID) dataset, but are not used often. Only 13 percent of existing DWH users has ever accessed the SID data. Barriers to using these data may include lack of understanding of ICD-9 codes and the inability to determine what codes are relevant to the chronic condition the user is examining. IDPH needs to explore why so few users access the SID data. It is possible that existing users are finding what is needed through the county health snapshot- the most frequently viewed, non-statistical feature of the warehouse. IDPH may also consider expansion of the indicators in the county health snapshots to include more chronic conditions of interest (e.g., diabetes).

**Socioeconomic and health insurance data** are utilized by approximately 40 percent of IDPH staff and DWH stakeholders (Figure 1). About 13 percent of average users indicate a need to access these data (Figure 2). Sources for these data are being examined for candidate datasets for the DWH.

---

---

All datasets discussed as a medium-level priority, except Medicaid and acute disease, should be made available in the DWH as flat, de-identified files. These data files typically serve as base files for use in calculating rates, indicators and other measures of health. Posting these data will take minimal effort, but provide potentially significant benefit.

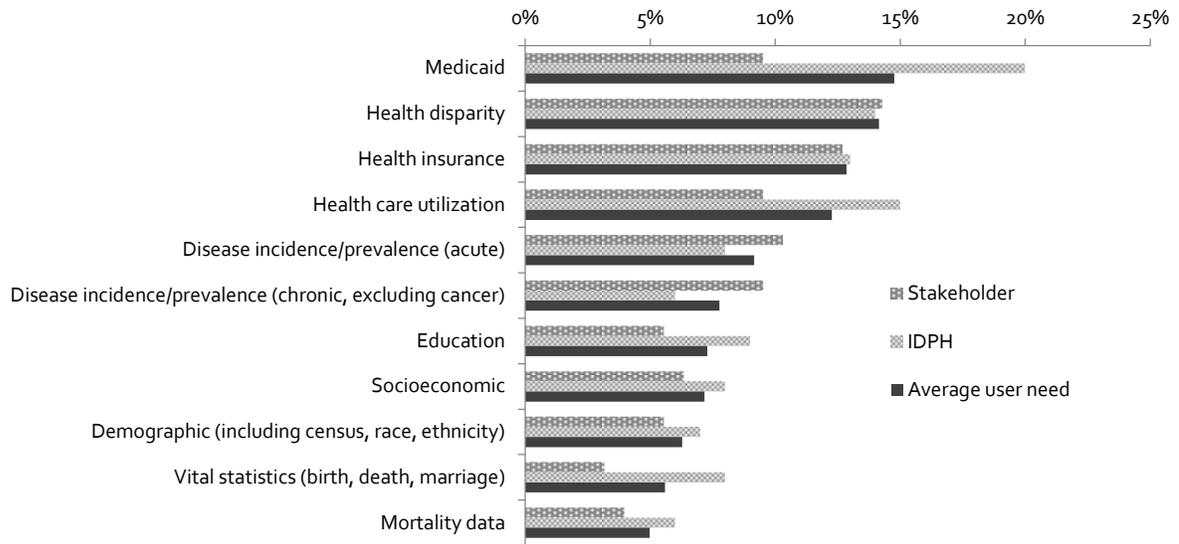


Figure 2. Percentage of data users need access to as reported by IDPH staff, stakeholders and average users



## Confidentiality requirements and data suppression

### High level priorities

Results of the assessment strongly indicate a need to address uniform, department-wide data confidentiality and suppression of data for the purpose of release. The majority of IDPH programs have confidentiality requirements for the release of data, but there is no uniform, comprehensive policy for the department. Of those programs with requirements, only 58 percent were reviewed by the Iowa Attorney General’s office (Table 2).

The department is working to develop standards, and potentially policy, surrounding data release and confidentiality. In order to be relevant and applicable to the DWH project, such standards must address the following:

- Release of all departmental data, not just data with known confidentiality issues such as acute disease
- Duplication in data release by the department (i.e., an assessment of other releases of the same data should be conducted to ensure lack of duplication and conflicts in statistics released from the same data source)
- Options for suppressing data (e.g., limited combination dataset display, review of existing reports catalog)
- Restrictions in data display for DWH users by user level

**Table 2. Percentage of IDPH programs with confidentiality policies, and those reviewed by the Iowa Attorney General**

Does your program have confidentiality requirements for the release of data?	Count	Percent
Yes	31	78%
(blank)	4	10%
No, we do not have requirements but release data	3	8%
No, we do not have requirements because we do not release data	2	5%
Were the confidentiality requirements for release of data reviewed by the Iowa Attorney General's office?		
Yes	23	58%
Unknown	9	23%
(blank)	7	18%
No	1	3%

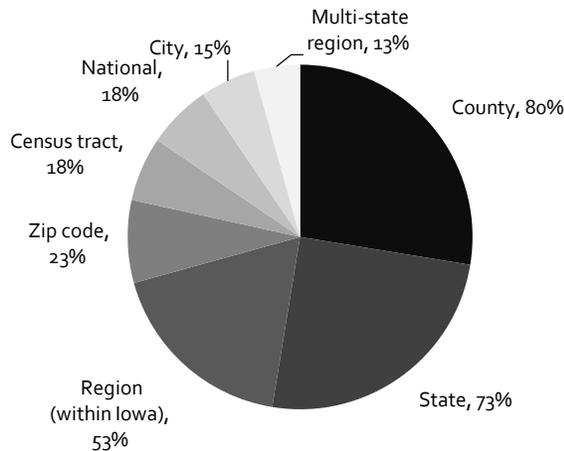
Business requirements must accommodate confidentiality requirements drafted by IDPH in the suppression and display of data in the DWH.

For the purposes of EPHT, the expectation is that states make available as minimally aggregated data as possible while still ensuring the confidentiality of all data in accordance with IDPH policies. The DWH must be able to dynamically suppress data cells containing small numbers on the public portal while still allowing authorized users on the secure portal to access unsuppressed data.

### Medium level priorities

Though the data are not published in this report, the assessment revealed that only 61% of IDPH programs have an editorial review process for publicly-released reports and only 49% have a method for performing a subject matter or statistical review. This and several other data governance-related issues are being cataloged and detailed by a Data Governance workgroup. The group will be making recommendations to department administration in the coming year.

## Display and visualization



**Figure 3. Percentage of geographic display level preferred, as reported by average DWH users**

and recommendations document is available as Appendix A.

Secure and public portals must provide a means to generate visual displays of data, including maps, charts, tables, and graphs. To obtain data on the environment, exposure, and health effects in both secure and public portals, states must provide ability for users to query the data. At a minimum, both portals must allow users to build and execute custom queries related to the standards of the EPHT program, including the following:

- Query display options – displaying the appropriate geography and demographic options for datasets selected in the query tool by the user.
- Ability to build a custom query – generates a custom query based on the options selected by the user.
- Query execution – the query tool executes the user-created query and generates output.

Query results should be represented in a table, and users must have the option to view data as a basic graph, charts, or static maps. Maps must display both counts and rates, provided no confidentiality conflict results. Visualization tools across the network will, for color sections, use standard color pallets from proven scientific research. Compliance with the American Disabilities Act (ADA) must be part of the business requirements. Specifically, ADA Section 508 standards detail color patterns, font size, and mouse-over functionality.

Additional resources and guidance on portal design and visualization from the EPHT Program are available in Appendix B.

When data are visualized in map, graph or chart format, contextual information, or footnotes, derived from descriptive metadata should be included in the display. This has been accomplished on other EPHT Network portals through the inclusion of query notes

For public display, there should be various ways to disseminate information including, but not limited to, Web-based query results and standardized reports. For advanced and basic users on the secure portal, data display may include Web-based query results, standardized reports, and the ability to export data files. Advanced users may potentially

### High-level priorities

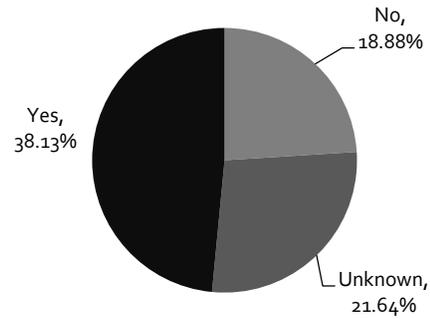
According to the IDPH and stakeholder assessments, data display at the state and county level are the most important (Figure 3). County and state level data display are the existing standard in the warehouse.

To maintain consistency across the national EPHT network, individual state portals will need to implement analysis, visualization and reporting formats and tools consistent with the National EPHT Portal standards. The complete standards

export data files with different suppression rules applied. User roles will be discussed in greater detail in the “User Roles” section of this report.

*Comparability of data*

Users have indicated a medium-level need to view more than one dataset at a time (Figure 4). Users should be able to choose the variables from at least two datasets to merge into one data display. This may include exporting data into a single excel workbook, selecting data display options in pivot table, selecting variables for display in a map and accompanying table, etc. The most common use of this capability is the ability to select counties, and view a dataset in common for those counties.



**Figure 4. Average users indicating whether there is a need to bring together multiple datasets in the DWH**

For basic users, any combination data display must be limited to suppression rules. The specifics of suppression rules are being discussed by the department’s Data Governance workgroup.

A core objective of the EPHT program is that analysis, visualization and reporting (AVR) tools must provide the capability to combine appropriate multiple data sources into a single graph, table, chart, or map.

Users have indicated a need to display data in the form of descriptive statistics (Table 3). Users should be able to perform basic calculations within the same data file the data is displayed in (e.g., currently that is an excel workbook). Users should be able to calculate summaries, averages, standard deviations and create a high-low range. These abilities are all standard in analytic programs such as excel.

**Table 3. Percent of analytic competencies used by IDPH programs**

Analytic competencies and outcomes used by programs	n	Percent
Descriptive statistics (sum, mean, standard deviation, range)	30	75%
Cross-tabulation (2x2 tables, risk ratio, odds ratio, stratification)	22	55%
Descriptive analytic functions (frequencies)	25	63%
Rates per population	22	55%
Trends	20	50%
Upper/lower limits	14	35%

*Additional display requirements*

Additional display requirements pertain to only EPHT datasets and indicators. These areas are currently under development:

- Landing pages for each content area will be developed to provide background information, supporting documentation, and links to additional resources related to each content area.
- For EPHT indicators, profiles should be available and provide source, derivation, rationale, strengths, limitations, and restrictions of use for the measures available on the portal.

**Medium level priorities**

The department should explore the ability to establish set regions of public health for comparison purposes.

An optional function is the capability to display data for multiple counties in one view. This option should include the ability for the users to select the counties he or she wants to view up to the maximum number limited by the display space.

*Display, format and manipulation of data*

Users indicate that the DWH is moderately easy to use, though there are features that could be improved upon (Table 4, Table 5). The ability to expand/contract the level of detail and calculate rates using the files in the warehouse is challenging for more than one quarter of existing users. However, the majority of users (67 percent) use Excel as their analytic program of choice (Figure 5). Data display within Excel should be examined to enable users to calculate rates, expand/contract a less-limited number of fields, and allow the creation of graphs.

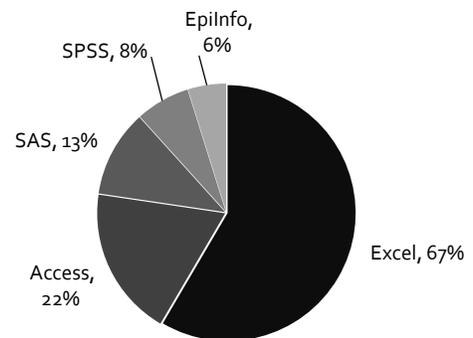
**Table 4. Stakeholder rankings of existing functions of the DWH**

How would you rate the following functions of the Data Warehouse?	n	Easy	Difficult
Filtering data	48	83%	17%
Sorting data	48	83%	17%
County-level view	47	83%	17%
State-level view	46	83%	17%
Ability to expand the level of detail	44	75%	25%
Ability to contract the level of detail	43	74%	26%
Rate calculations	43	72%	28%

**Table 5. Stakeholder rating of the ease of the overall use of the DWH**

How would you rate the Data Warehouse in terms of ease of use of data?	n	Percent
Moderately easy to use	38	30%
Difficult to use	8	6%
Extremely easy to use	2	2%

**Figure 5. Distribution of frequently used analytic programs as reported by average DWH users**





## User roles

---

### High level priorities

In its current state, DWH users may access the application through either a direct Analysis Services data connection via Excel or an Excel Services connection via a SharePoint site. SharePoint access (basic user) requires Active Directory authentication. The direct Analysis Services connection is available to only advanced users through the IDPH network. In order to meet the requirements of the EPHT program, modifications must be made to the DWH to provide distinct secure and public access points. Both access points should be available via the web, and secure access must additionally comply with State of Iowa security standards. Currently, advanced and basic users both have the ability to view and export unsuppressed data, and those abilities should be retained.

The existing DWH user roles do not meet the EPHT requirement for a public user role with access to suppressed data. The detailed needs of all system users should be gathered during the requirements process and organized into logical user roles. An overview of the existing roles and proposed public role are outlined below.

#### *Advanced user*

Currently advanced users must have a direct connection to the IDPH servers to access data in the warehouse. In order for advanced users to access the data external to the department, a new security and authentication solution must be identified. The current solution requires token use to access the data through a VPN connection, which is cumbersome for users. The use of tokens is also costly and not feasible for advanced users who are not department employees. The final solution set should explore a solution with improved user experience and reduced cost.

Stakeholders have detailed which advanced user functions are important. Advanced users have the ability to select options for a table view through the use of pivot tables in Excel. However, users do not have the ability to create county-level maps with one or more datasets. Solutions for mapping data in the warehouse should be reviewed in the business requirements writing phase. About 20 percent of stakeholders express a need to view detailed data in a secure environment, with the recognition that the detailed view of data would not be available at other user levels.

Stakeholders communicated a need to access patient-level data for the purpose of public health research. IDPH uses a review committee, similar to an institutional review board, to approve or deny access to department data. Researchers may request patient-level data, with identifiers, as long as the research has a public health focus. In addition, 27 percent of stakeholders indicated a need to access patient-level data now or in the future. A solution for storing patient-level data should be explored, but is not a high-level need at this time. It will be critical to consider how this solution might be maintained, how approval for access will be obtained, how external access might be accomplished, and whether there is potential for revenue-generating in the solution.

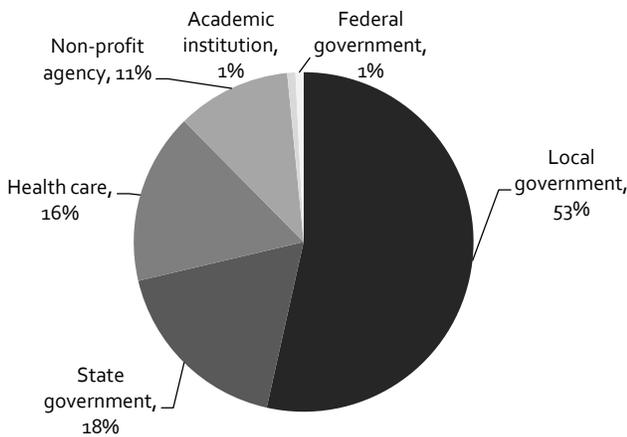
#### *Basic user*

The majority of stakeholders accessing the warehouse outside of the department (62 percent) will not need to view detailed, patient-level data (Table 6). However, there is a need to retain a user role with access to unsuppressed, county-level data, but with different permissions than advanced level users. Currently, basic users have the ability to see and export unsuppressed data, but have limited capability to manipulate the data. This user level is valued by local public health partners who constitute more than 50 percent of users (Figure 6).

---

**Table 6. Preferred options for advanced users, research and patient-level data access as reported by stakeholders**

<b>If an advanced user level were available through the IDPH Data Warehouse, which functionalities would be important to you (check all that apply):</b>	<b>n</b>	<b>Percent</b>
Ability to create county-level maps with one or more datasets	70	56%
Ability to select options for a table view	47	37%
Secure access to detailed data (individual level)	28	22%
Ability to view data suppressed in other user levels	27	21%
Other (please specify)	2	2%
<b>Do you have a need to access patient-level data for the purpose of public health research?</b>		
No	78	62%
Yes	24	19%
(blank)	24	19%
<b>In your current position, do you have or anticipate a need to access patient-level data housed at the Iowa Department of Public Health (e.g., birth, death, marriage, chronic disease data)?</b>		
(blank)	41	33%
Yes	34	27%
No	33	26%
Unknown	18	14%



*Public user*

This user level is a requirement of the EPHT project. According to CDC standards, Nationally Consistent Data Measures (NCDMs) must be made available through a publicly-accessible portal. The definition of "portal" is flexible. Precedence set in other states with EPHT programs vary in public data display from posting PDF files with NCDMs, to interactive portals with mapping, tables, graphs and reports. It is an expectation that all EPHT states' portals eventual have interactive capabilities

**Figure 6. Employer categories reported by stakeholders**

Suppression of confidential data varies from state to state. The Iowa standard will be to display no fewer than five events per county for a set time parameter. Detailed suppression rules are under development. Possible solutions for a public portal and its functionality should encompass highly interactive functionality, as described in the "Display and Visualization" section of this report, down to simple reports displaying NCDMs and other health indicators. Detailed public user requirements should be gathered and documented during the requirements process. IDPH anticipates reviewing proposed solutions before inclusion into the business requirements.

It is not known whether there is a need to distinguish between "Basic" and "Public" user roles. Possibilities for separate or shared roles should be included in the business requirements.

*Additional securities and permission levels*

At present, the DWH has four permission levels- administrator, business administrator, advanced user, and basic user. The distinction between levels is based on ability to alter the contents and structure of the DWH, manage users and display



data. Administrators have full-control over the application. Business administrators may manage users, post documents and view data at any level.

Current permissions should be compared to EPHT standards, and may need to be redesigned with the addition of a "Public" user role.



## Data processing, ownership policies, data acquisition and migration timing

---

There are two roles used in the management of the data in the DWH. Data owners are the data originators, or persons responsible for providing the data to IDPH Information Management. Data stewards are responsible for testing and reviewing data in the test environment to assure accuracy when compared to the original dataset and published forms of the same data. DWH new dataset inclusion standards require that a data owner and steward be identified. With some datasets, those roles are synonymous.

There may be a need for EPHT data to be processed by the data owner and/or IDPH prior to loading into the DWH. Processing (aggregation, recoding of variables, pre-calculation of rates, etc.) will vary depending on the nature of the dataset.

The DWH must also be capable of exporting data tables for submission to CDC for use on the National Environmental Public Health Tracking portal. Datasets unrelated to the EPHT program and stored in the DWH do not need to be transmitted to an external agency at this time.

### High level priorities

#### **Policy for establishing Data Sharing Agreements with entities external to IDPH (excludes research agreements)**

IDPH has a standard agreement for data exchange between the department and external agencies. All data sharing agreements must be reviewed by the Iowa Attorney General's office, as well as the legal staff from the external agency. When initiating a new agreement, IDPH staff must assess whether an agreement exists and if the new data access needs may be covered by an existing agreement.

Inquiries for access or use of data outside the functions of the public and secure portals will be directed to the data owner for review and approval.

#### **Policy for transfer of data into the department from external entities**

It is imperative that data acquired for use in the DWH be transferred to the department through the Bureau of Information Management. Initial transmissions of data may be facilitated by the requesting program, but a schedule of routine transmission must be established directly from the external entity to the Bureau of Information Management before data is integrated into the warehouse.

#### **Proposed standards for new dataset inclusion into the Data Warehouse:**

- 1) Data must be available for the years starting no earlier than 2000 through the present year. Datasets originating after 2000 may be considered for inclusion into the DWH.
  - 2) Support for continued collection of this data in future years must be identified.
  - 3) Datasets must have an identified data owner that is an individual, not an agency. Each dataset must also have a data steward committed to testing the validity of the data once migrated to the test environment. The steward also serves as the consultant to IDPH when questions about the dataset arise. In some instances, these roles are held by the same person. The data steward is also responsible for facilitating the development and updating of metadata.
  - 4) Key variables for the dataset must be reasonably static from year to year. Data dictionaries for each year must reflect changes in key variables; or the data owner must provide a detailed description of changes in key variables including, but not limited to:
-



- a. Changes in field names
  - b. Changes in field types (e.g., numeric to text)
  - c. Deletion or addition of fields
  - d. Changes in coding for field responses
  - e. Other changes affecting response-level data
- 5) Data owners must be able to provide all information requested in the IDPH Dataset Catalog
  - 6) Data owners should provide a list of entities that request and receive data with the intent to publish. This includes the names of entities that receive the data for mandatory reporting for grants, legal mandates or the like.
  - 7) Data owners must commit to a set annual release date for the data. Modifications to the data made after the annual release date must be documented in DWH metadata.

*Annual load schedule*

The DWH is a repository that hosts static data. Integrity and the absence of historical changes to the data are imperative. The annual data load schedule is dependent on the release of census estimate files containing county-level population counts with race/ethnicity categories and single year age. The following is a proposed schedule for annual data load:

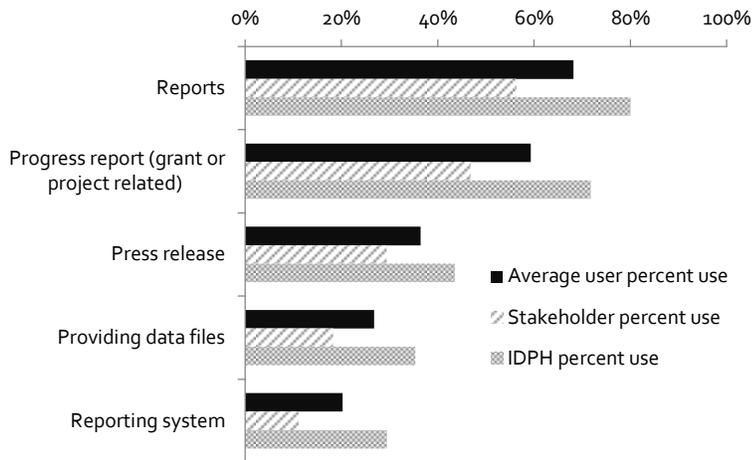
**Table 7. Annual load schedule for the DWH**

Dataset	Month of typical release	Proposed month of migration to the test environment	Proposed month of migration to the production environment	Required month of data transmission to CDC*
Census	September			
Birth	July	August	September	
Death	July	August	September	
BRFSS	June	August	September	
Statewide inpatient discharge (SID)	April	August	September	October**
Lead- child	April	September	October	May
Water quality	May	September	October	May
Air quality	May	September	October	
Cancer	June	September	October	
Birth defects	June	September	October	October

\*Requirements of the EPHT program include bi-annual transmission of data to CDC.

\*\*Data reported to the CDC from SID include only events of hospitalization for asthma, myocardial infarction, and carbon monoxide poisoning.

## Reporting



**Figure 7. Percentage of formats for release of data as reported by IDPH programs, stakeholders and average users**

### High-level priorities

#### *Data release*

The average DWH user disseminates data in report format. A high percentage of IDPH staff (72 percent) release data in grant or project-related progress reports (Figure 7). As stated previously, users of the data in the warehouse must be able to calculate simple descriptive statistics such as rates, averages, and standard deviations and should also be able to create graphs, charts and tables from the data in the advanced user view.

#### *County health snapshots*

Of all non-statistical features in the DWH, the county health snapshots are used most often. These snapshots and data in the warehouse are used by local public health agencies in conducting health assessments, creating indicators, and examining trends in their populations (Table 10). On average, indicators are used by 44 percent of IDPH staff and DWH stakeholders (Table 8).

**Table 8. Use of data in monitoring population health, reported by average users**

Do you use data to monitor the health of a population?	Average
Yes	44%
No	28%
Unknown	8%

IDPH, with the help of external users, need to re-examine indicators in the county health snapshots. Indicators should be reviewed while business requirements are being written. IDPH staff should involve the program leads for Healthy Iowans, CHNA-HIP, and external user groups. NCDMs from the EPHT program should be added to the snapshots. Generation of the health snapshots is automated, and should remain automated.

Less than 10 percent of users access supportive tools such as the Glossary or Toxicology Manual (Table 9). IDPH staff may decide to evaluate the usefulness of having these documents in the warehouse.

#### *Supportive documentation*

Documentation is a tool that provides valuable assistance to users. To assist in navigation, an information system should provide strong documentation. Such documentation can include instructional use on navigating the portal, explanatory text on data interpretation, and public health implications. Documentation can vary depending on user expertise; thus the challenge is to provide technical documentation to assist the expert user while at the same time providing general documentation to help non-expert and general public users.



Which non-statistical features have you ever accessed in the Data Warehouse?	n	Percent
County health snapshot	39	31%
Data Warehouse Training Materials	18	14%
Epi Manual	12	10%
Metadata	12	10%
Foodborne Outbreak Investigation Manual	11	9%
Environmental Health Toolkit	10	8%
Glossary	9	7%
Toxicology Manual	3	2%
Intro to ArcGIS for IDPH	2	2%
Using Adobe Acrobat to Create Custom Forms	1	1%
ATR Provider Resources	1	1%

**Table 9. Use of non-statistical features as reported by stakeholders**

Common questions can arise across any system. Developing a set of Frequently Asked Questions (FAQ) will help reduce the burden of answering most of these common questions. As new questions are identified, the FAQ should be updated. A glossary of commonly used and clearly defined terms should complement the FAQ, and definitions should be included in the online documentation. Additional details may also be included in the online documentation.

Users may have questions not addressed in any documentation. When possible, additional information will help clarify issues and answer user questions. State portals should provide a mechanism for users to ask questions and request clarification. This mechanism may be in the form of an email address or phone number.

**Table 10. Type of DWH use as reported by stakeholders**

How have you used the Data Warehouse data over the past year (check all that apply)?	n	Percent
Complete a community health needs assessment (CHNA)	28	22%
Assess health indicators for a county	26	21%
Locate statistics on health conditions	24	19%
Complete a county health improvement plan (HIP)	23	18%
Examine health trends in a population	19	15%
For a grant application	13	10%
Rates or counts on a specific condition	12	10%
Increase general knowledge of health data	10	8%
To respond to an external request for data	9	7%
For a publicly-released report with data content	6	5%
Form health indicators	6	5%
Explore the data warehouse	4	3%
To respond to a media inquiry	1	1%



## Metadata standards

---

### *Current Metadata State*

Metadata were initially collected from subject-matter experts (SME) at IDPH. Metadata were created and collected using Microsoft Word and then input into the Department's Universal Help system which is used as the Department's primary tool for reference materials and help files. The metadata are displayed on the home page of the basic user site and can be navigated from there. Advanced users can access the metadata via the basic site or using the public URL

[http://www.idph.state.ia.us/idph\\_universalhelp/Main.aspx?System=IdphDataWarehouseMetadata](http://www.idph.state.ia.us/idph_universalhelp/Main.aspx?System=IdphDataWarehouseMetadata).

Users can view metadata by subject area (i.e. births, deaths) which includes the following elements: subject area overview, data collection methods, reporting suppression criteria, limitations, disclaimers, additional resources and the IDPH point of contact. Within each subject area, users can drill into the available data attributes in the warehouse and view the following metadata elements: attribute definition, purpose, limitations, disclaimers, calculation methods, and other notes.

Additional metadata are captured for the County Health Snapshot including an overview of the report, purpose of the report, limitations, disclaimers and suppression criteria for measures contained in the report. At this time, metadata are not searchable.

Review and revision of metadata are the responsibility of the IDPH subject area contact as listed in the metadata. Word documents containing the metadata are stored on a shared drive and accessible at all times to the contacts for revision. The IM project manager currently reaches out to each subject area contact prior to moving a new set of data into Production to request the metadata be reviewed. If there are subsequent revisions made to the Word document, the IM project manager then updates the metadata content in Universal Help. This is an entirely manual process.

### **High level priorities**

#### *Desired Future Metadata State*

The future state of metadata for the DWH is not yet fully defined. While the requirements of the EPHT portal are known, IDPH has not yet fully explored or consequently defined metadata requirements. As such, the EPHT metadata requirements will be described at a high level in this document and we anticipate that as part of the DAS-ITE requirements engagement, overall IDPH metadata requirements would be refined and documented.

The Department strongly desires to build upon the EPHT metadata requirements and avoid duplication of efforts in collecting and maintaining metadata. While the EPHT team explores several free applications for inputting metadata and generating XML files, there is yet to be defined roles, responsibilities or processes for the Department's collection and maintenance of metadata. This topic is anticipated to be explored in the future by the IDPH Data Governance workgroup.

#### *EPHT Metadata Requirements*

The EPHT metadata standard is based on the existing Federal Geography standard, but only requires a subset of information be provided. This standard assumes the data is geographic in nature which may not apply to all datasets contained in the DWH. Metadata must be made available to all consumers of EPHT measures, regardless of role. Additionally, users must be able to search metadata.

Metadata should be incorporated into the DWH in a way that allows it to be consumable in multiple ways. The public portal for the DWH should provide for the ability to search metadata.

---



Grant requirements for the EPHT program include submission of metadata in XML format to CDC for storage in the National Metadata Repository. To comply with the grant requirements, IDPH is responsible for generating valid XML files for submission to the CDC twice a year. These files must follow the CDC metadata model (Appendix C).

The [EPHT metadata field guide](#) provides the EPHT required fields for submission to the CDC.



---

## Financial sustainability

---

### *Sources of continuous funding*

The Environmental Public Health Tracking grant is a driving force behind the expansion and enhancement of the DWH. The current project period lasts until July 31, 2014. During the project period IDPH must demonstrate progress and meet certain benchmarks in the development of an Iowa Tracking Network to receive annual funding. At the end of the current project period it is anticipated that CDC will solicit competitive applications for continued funding.

While it is anticipated that Iowa will have annual funding to support EPHT and the DWH through July 2014, the current federal budget environment is such that funding through the EPHT program may not be guaranteed beyond the current budget period ending July 31, 2012.

Part of the business requirements writing process should include a review similar data systems and revenue generation. A handful of states with functional data warehouses or EPHT portals have the ability to charge a fee for user access to data. Potential limitations to implementing a fee-for-use component to the DWH are twofold; 1) much of the data available in the DWH is accessible to the public per the Iowa Administrative Code, 2) IDPH may not have the infrastructure to manage a fee-based system. At a minimum, IDPH and DAS/ITE should determine whether fees may be levied for academic researchers who request large datasets. These requests often take a significant amount of staff time to prepare the data for release. Non-governmental agencies routinely charge for access to data and such fee structures should also be considered.

### *Funding sources for other state agency DWH applications*

DAS/ITE will be asked to review sources of financial support for other data warehouse applications in other state agencies.

