From Words to Concepts to Queries:

Helping Users Find Series and Variables

To Satisfy their Information Needs

Final Report to the United States Bureau of Labor Statistics

Stephanie W. Haas, Ph.D.

stephani@ils.unc.edu
School of Information and Library Science
University of North Carolina at Chapel Hill

November 12, 2001

Executive Summary

One of the problems facing members of the public when they seek information from the United States Bureau of Labor Statistics website (http://www.bls.gov) is the gap between the words and concepts that they use and understand, and the terms and concepts that the BLS uses, as expressed in its surveys, series, and other publications. The goal of this work is to explore ways of providing information to help users bridge the gaps between their questions and the answers provided by BLS information. It focuses on the gaps between end user words and concepts and BLS terms and definitions, although related problems are also addressed. It builds on the earlier development of the LABSTAT Crosswalk (LSC) (Haas, 2000) and the work on development of a query grammar by Liddy & Liddy (2001). For the purposes of this research, I concentrated on two major concepts, *industry* and *occupation*, along with some related concepts. There are three major components of this work.

1. The **LABSTAT Crosswalk (LSC)** (Haas, 2000) is a reference resource that maps common, general language words and phrases for BLS concepts to the agency terms, especially those used in the LABSTAT database via either Selective Access or Public Data Query (http://www.bls.gov/sahome.htm). Liddy & Liddy (2001) analyzed a corpus of user queries to develop a query grammar that describes common question structures, along with lists of words and phrases that can fill each slot in the grammar. The major categories in the grammar were grouped according to "who, when, where," etc. as components of a query. Harmonization of these two schemes showed that the LSC clusters provide broader coverage than the query grammar, however, the differences are straightforward, in part because the LSC clusters are generally finer grained than the grammar categories.

2. The **Matrix Model of Survey/Series and Variables** is based on the premise that an important pre-requisite for identifying a good source for the answer to a question is an understanding of what the available sources are, and what they contain. The rows of the Matrix list the clusters, categories, or variables by which BLS information can be specified, e.g., *race*, *gender*, *occupation*, *industry*, and so on. The columns list each survey, table, or series. The Matrix could be developed into a directly manipulable interface that allows users to explore their choices, or a background data store that provides information for other types of interface tools. It could support a variety of tools to help users formulate their queries, or to help them diagnose problems in their queries.

3. The **Progression Model** shows the progression from user question to query formulation, incorporating the different transformations the question may have to undergo, as well as the types of information aids or tools that might be helpful at each step. This model follows a traditional pattern of information seeking; once the user has a question, he/she must select the best source(s), and formulate the question in such a way that it can be answered from the source. In a traditional library, reference librarians are immediately available to provide assistance. When using a website, however, intermediaries may be available only by email. Each translation or

transformation the user must make, from his/her own vocabulary to BLS concepts, from BLS concepts to relevant survey/series, from relevant survey/series to the variable-value pairs that form a query, can be eased if there are a variety of aids, information tools, reference sources, etc. available for use. This model claims that the three crucial points at which users may need additional information in order to build a successful query are at the translation or transformation points. To that end, it proposes three types of information aids: **word – concept aids**, **concept – survey/series aids**, and **survey/series – query aids**. Aids can range from simple definitions or examples to automatic parsers or focused browsing interfaces. Four extended examples of how the Progression Model can help organize and present helpful information to end users are shown in Appendix A of the report.

The work described in this report has explored ways of improving end user access to BLS information, with two main emphases.
1. How to bridge the gap between users' concepts and vocabulary and the concepts and terms used in the BLS domain.
2. How to provide information to the users at each stage in the search process that will inform them about the information available, and enable them to make good decisions in formulating queries. This process includes identifying the BLS concepts that correspond to the concepts in their queries, selecting survey/series that are relevant to these concepts, and selecting variable-value pairs.

The two models that were created during the course of this work illustrate where the crucial decision points are in the user's information-seeking process, and demonstrate some means by which useful information could be provided.

1. To Bridge Between User Vocabulary and BLS Variables and Values

One of the problems facing members of the public when they seek information from the United States Bureau of Labor Statistics website (http://www.bls.gov) is the gap between the words and concepts that they use and understand, and the terms and concepts that the BLS uses, as expressed in its surveys, series, and other publications. For example, someone who needs information on employee compensation in small businesses has several of these gaps to cross.

- What BLS concepts and terms correspond to the idea expressed by *small business*?
- Is the BLS definition of *small* consistent with the user's notion?
- What surveys or series provide information on employee compensation broken out by size of business?
- Once the correct surveys/series are found, what are the names of the relevant variable(s)?
- Once the relevant variable(s) are identified, what value(s) should the user select?
- Are there any interactions between variables that might limit the user's choice of values?

This user's problem is an instance of the traditional reference problem. Thinking of the question is merely the first step of the process. Subsequent steps require identifying the appropriate information resources, and then expressing the question so that the best available answer will be returned. An interaction between the user and the BLS information resources requires the user to make a series of decisions. His/her decisions are more likely to lead to a useful answer if the user has some knowledge about the way in which information is provided.

The goal of this work is to explore ways of providing information to help users bridge the gaps between their questions and the answers provided by BLS information. It focuses on the gaps between end user words and concepts and BLS terms and definitions, although related problems are also addressed. It builds on the earlier development of the LABSTAT Crosswalk (LSC) (Haas, 2000) and the work on development of a query grammar by Liddy & Liddy (2001). For the purposes of this research, I concentrated on two major concepts, *industry* and *occupation*, along with some related concepts. This report includes:

- Discussion of the harmonization of LSC and Liddy & Liddy grammar categories.
- Matrix Model of Survey/Series[1] and Variables, first presented in the May 14, 2001 progress report, along with ideas for its use in supporting user queries.
- Model of the Progression from Question to Query Formulation, which suggests information and reference aids that could be useful at each stage.
- Four extended examples of how the Progression Model can help organize and present helpful information to end users. (Appendix A)
- Background information showing the LABSTAT surveys/series and variables in which the *industry* and *occupation* concepts occur. (Appendix B)

---

[1] Throughout this report, the term "survey/series" is used as shorthand for surveys, tables, time series, regular publications, special papers, and any other formats in which the BLS publishes data.

- Background information mapping end user words and phrases associated with the *industry* and *occupation* concepts to BLS surveys/series and variables. (Appendix C)

2. Harmonization of LSC and Liddy & Liddy grammar categories

The LABSTAT Crosswalk (LSC) (Haas, 2000) is a reference resource that maps common, general language words and phrases for BLS concepts to the agency terms, especially those used in the LABSTAT database via either Selective Access or Public Data Query (http://www.bls.gov/sahome.htm). The LSC information structure organizes the general words and phrases in which end users describe the concepts that are components of their query into semantic clusters, and then maps them to BLS terms and the information products (e.g., surveys, tables, series) in which they can be found. Examples of semantic clusters are *industry*, *ownership*, *establishment size*, and *occupation*. General language words and phrases were gathered through environmental scanning of a variety of sources such as newspapers, magazines, radio broadcasts, and textbooks.

Liddy & Liddy (2001) analyzed a corpus of user queries to develop a query grammar that describes common question structures, along with lists of words and phrases that can fill each slot in the grammar. The major categories in the grammar were grouped according to "who, when, where," etc. as components of a query.

These two analyses of general language describing labor-related concepts thus originate from slightly different sources. The LCS clusters were primarily drawn from informative sources from which people might learn or be reminded of relevant words and concepts. The Liddy & Liddy grammar was based on actual user queries to the BLS. This was therefore a more restricted language corpus, however, it was representative of how people phrase their questions. These two methods of gathering user words and phrases serve complementary functions. The data from user queries show words and concepts that people use frequently, which can thus be considered as priority topics for development of user aids. Environmental scanning provides a broader view of the context in which people search for information, and can thus also indicate the types of information related to core topics that could help satisfy users' information needs.

The goal of harmonizing these two analyses was to determine if there were any significant differences in the concepts or expression of the concepts. As expected, the LSC clusters provide broader coverage than the query grammar, however, as Table 1 shows, the differences are straightforward, in part because the LSC clusters are generally finer grained than the grammar categories.

LSC clusters	Liddy & Liddy categories
Absent from work	What – employment
Age	Who – age
Benefits	What – general
Consumer costs	What – economic indicators
Data products	No equivalent

Disability	No equivalent
Diversity	No equivalent
Economy	What – economic indicators
Education	What – education
Employment	What – employment
Ethnicity	Who - ethnicity
Family status	Who – household
Gender	Who – gender
Geographical area	Where
Hours	What – employment
Income	What – general
Industry description	No equivalent
Injury	No equivalent
Occupation	What – occupation
Job tenure	What – employment
Labor force	What – employment
Legal	No equivalent
Marital status	Who – marital status
Other demographic	No equivalent
Population	What – employment
Producer costs	What – economic indicators
Race	Who – race
Statistics	No equivalent
Time	When – time
Type of industry	No equivalent
Unemployment	What – employment, economic indicators
Unions	No equivalent
Units	Who – household
Wage, salary, income	What – general
Work arrangements	No equivalent

Table 1. Correspondence between LSC clusters and Liddy & Liddy categories.

These semantic clusters/categories, along with the extensive list of words and phrases that belong to each, represent end users' concepts and vocabulary. Individual users have greater or lesser amounts of expertise and familiarity with BLS terms and definitions, and therefore greater or lesser understanding of how to "translate" their questions into queries that can retrieve useful information from BLS sources. Having identified a representative set of concepts from the users' perspective, the next step was to start identifying equivalent concepts from the BLS perspective.

3. Matrix Model of Survey/Series and Variables

The Matrix Model of Survey/Series and Variables is based on the premise that an important pre-requisite for identifying a good source for the answer to a question is an understanding of what the available sources are, and what they contain. For example, if a question concerns the *industry* concept, any survey/series that does not include an industry-related variable can immediately be eliminated from consideration. The difficulty is that the BLS has a rich collection of information resources; even focusing only on those available through Selective Access and/or Public Data Query presents a

confusing choice for an end user. The Matrix serves as a single, organized catalogue of what concepts or variables are available in the various surveys/series. The BLS has prepared similar finding aids for specific topics, such as Wages by Area and Occupation (http://stats.bls.gov/blswage.htm). The Matrix could be developed into a directly manipulable interface that allows users to explore their choices, or a background data store that provides information for other types of interface tools. It could support a variety of tools to help users formulate their queries, or to help them diagnose problems in their queries.

The rows of the Matrix list the clusters, categories, or variables by which BLS information can be specified, e.g., *race*, *gender*, *occupation*, *industry*, and so on. The columns list each survey, table, or series. Figure 1 shows a portion of the Matrix. An **X** in a cell indicates that the survey, table, or series contains data organized by that category.

	Mass Layoff Statistics	**Producer Price Index**	**National Employment, Hours, and Earnings**
Seasonal adjustment	X	X	X
Sex	X		X
Age	X		
Race	X		
Geographical area	X		
Industry	X	X	X
Occupation			X

Figure 1: Portion of the Matrix Model of Survey/Series and Variables.

The rows of the Matrix can serve as the attachment point (literally or figuratively) to which we map templates of end user questions or their component concepts. The matrix structure itself (i.e., the column and row labels) is constructed from the LABSTAT data dictionary **Series ID Formats** (http://stats.bls.gov/hlpforma.htm). Since the series are generally stable, this should primarily be a one-time process, with updates required only as changes are made to LABSTAT.

Figure 2 gives an overview of how the Matrix could fit into user interactions with BLS resources. Construction of the Matrix is essentially a one-time process which organizes the information available in LABSTAT. There are three paths a user can follow to help translate his/her question into a query suitable for LABSTAT.
1. Work directly with the Matrix. The user may browse the Matrix to see what information is available where, or may focus on specifying or negotiating a query, or diagnosing why a query did not produce the expected results.
2. The user can consult any number of reference aids to gain a greater understanding of what information is available and how to best phrase a query, and then work directly with the Matrix as in the first path. Possible aids are discussed further in Section 4.

3. Sets of templates, or a query grammar (see Section 2) can parse the user's general
    language question, either automatically or in cooperation with the user, and map the
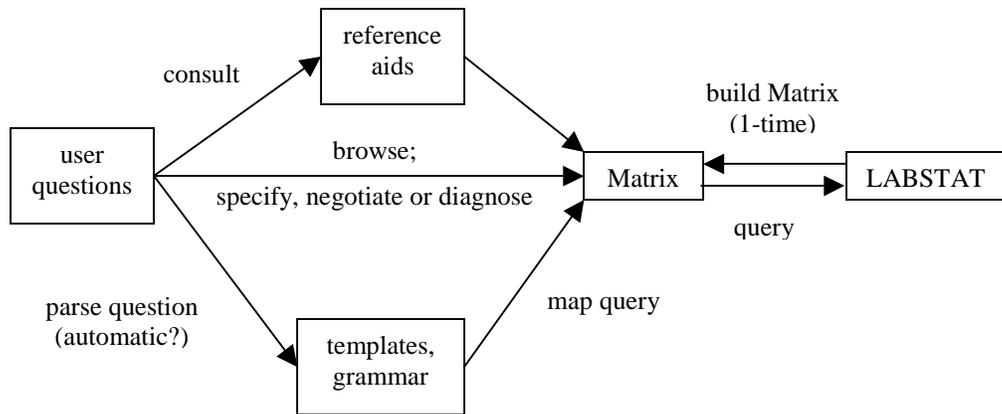   question components to the appropriate cells of the Matrix.

Figure 2. Context of use for the Matrix Model of Survey/Series Variables.

Possible Uses of the Matrix
1. A graphical browsing tool. Users can see what variables are available in a given
    series, or what series have data broken down by a given variable. This is similar to
    Marchionini's work on the relational table browser (Marchionini, 2000; Marchionini
    et al., 2000). Another form of browsing would allow a user to select, or "fix" a value
    for a variable, and explore the variables and values that then become available in the
    relevant survey/series, given the fixed value.

2. A dynamic aid for formulating queries.
    User selects table of interest.
    >   Unselected columns disappear or are grayed out.
    >   Irrelevant rows (variables that don't appear in the table) disappear or are grayed
    >   out.
    >   Relevant rows can expand in one of several ways:
    >   o   To show definition as used in that table
    >   o   To show values that can be selected (for simple variables such as <u>race</u>
    >   or <u>sex</u>)
    >   o   To show "groups" of values that can be selected (e.g., for complex
    >   values such as <u>industry</u>, these could include major industry sectors,
    >   such as *service* or *manufacturing*)
    >   o   To show related variables. E.g., in *State and Area Employment,*
    >   *Hours, and Earnings*, the *industry* concept is divided into two specific
    >   variables: <u>detail</u> and <u>industry</u>.

User selects category of interest, e.g., *industry*.
Unselected rows disappear or are grayed out.
Rows for related variables (e.g., <u>ownership</u>, <u>establishment size</u>) are added.
Columns that don't contain selected category disappear or are grayed out.
Rows could be expanded as described above.

3. An interface for entering queries.
After viewing possible choices, user clicks on specific values desired.

There are some remaining concerns about the development and deployment of the Matrix. Its development will require time and effort on the part of those familiar with the contents of BLS surveys/series. It would be another item requiring regular maintenance to ensure it accurately reflects the contents of the surveys/series.

In its fully developed and populated form, it will be a large, complex knowledge structure. Because of this, it may not be well suited for direct user interaction. One alternative is to devise some way of having users "drill down" into it, so that they never have to deal with the whole structure at once. Another possibility is to divide it into pieces, such that each piece represents a small set of related concepts, e.g., all those related to *occupations*, or to *compensation*. This would probably entail some redundancy between pieces, however. A third possibility, both for development and deployment, is to include only a limited set of concepts, e.g., those that users request the most, or that users find most difficult to work with.

4. Progression from Question to Query Formulation

In this section, we introduce a different model of the progression from user question to query formulation that incorporates the different transformations the question may have to undergo, as well as the types of information aids or tools that might be helpful at each step. First, we give an overview of the model (see Figure 3), which is followed by more details concerning each component.

The process starts when an end user comes to the BLS website with a question. The question is expressed in words and phrases with which the user is familiar. They may be general language words, or more specialized, technical terms (Haas & Hert, 2001). The first step is for the concepts represented by the user's vocabulary to be mapped to one or more BLS concepts. If this is impossible, it is an indication that the user's question is not in the scope of BLS information, and that another source may be more appropriate. A question may contain only a single concept, such as *unemployment*, but more often will contain more than one, such as *current unemployment* (which incorporates *time*) or *unemployment in North Carolina* (which incorporates *location*).

BLS concepts are those that are used in the surveys, tables, series, and other publications produced by the BLS. They are expressed using BLS terms (which may overlap to a

greater or lesser extent with those of general language (Haas & Hert, 2001)) which are used in a variety of places, including:

- survey question, e.g., asking if the respondent works **full time**,
- series titles, e.g., ***Nonfarm Payroll Statistics***,
- names of variables, e.g., **<u>industry</u>, <u>sector</u>,**
- values of variables, e.g., **blue collar occupations**, **small private sector establishments**.

A single concept may appear in many survey/series, and the user's task at this point is to identify the survey/series that contain both the target concept and the other concepts expressed in the original question. A concept such as *industry* appears in many places, but if the question concerns the number of layoffs in various industries, *Mass Layoff Statistics* is the only relevant series. Identifying the relevant survey/series frequently involves more than just looking at the title; variables, definitions, and other documentation may also be required.

Once the relevant survey/series have been identified, the user's task is to choose the variable-value pairs that best express the original question. Several factors can complicate this step. There may be variables that seem to have no direct relationship to the question, but for which values (at least a null value) must be chosen. There may be interactions between variables such that a value selected for one variable restricts the available choices for another. Sometimes the interaction may be obvious, as in the Selective Access sequence of choices for *Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)*. First, the user must choose a value for <u>detail</u>, e.g., *1-digit industry*, or *2-digit industry*, which restricts the list of values that will be displayed for <u>industry</u>. If the user's choice is *1-digit industry*, only the short list of 1-digit SIC codes will be displayed. This has the obvious benefit of shortening the list of values for <u>industry</u> through which the user must browse. In other cases, the interaction between variables is more opaque, especially to a novice user. A common problem is that selecting a more specific value for one variable forces a more general value for another. For example, in *Labor Force Statistics from the Current Population Survey*, if the user chooses a specific age range, such as *25-34*, there are no choices displayed for <u>occupation</u>. The more specific occupation choices such as *secretaries, stenographers & typists*, or *teachers, college & university*, are only available if <u>age</u> = *20 or older*, or is not limited at all.

This model follows a traditional pattern of information seeking; once the user has a question, he/she must select the best source(s), and formulate the question in a way that it can be answered from the source. In a traditional library, reference librarians are immediately available to provide assistance. When using a website, however, intermediaries may be available only by email. Each translation or transformation the user must make, from his/her own vocabulary to BLS concepts, from BLS concepts to relevant survey/series, from relevant survey/series to the variable-value pairs that form a query, can be eased if there are a variety of aids, information tools, reference sources, etc. available for use.
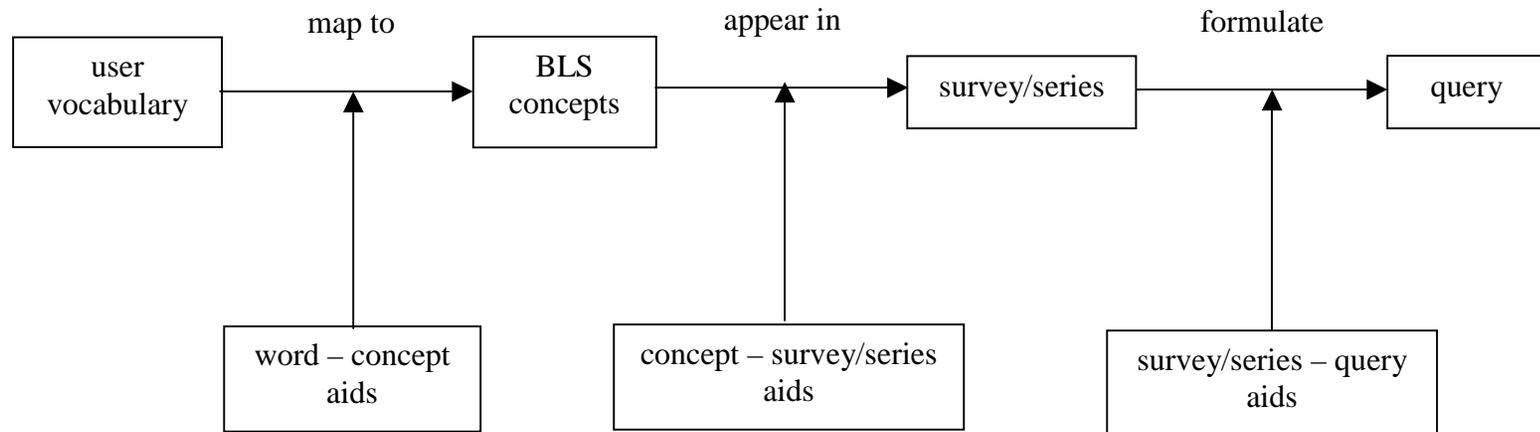
Figure 3. Overview showing the progression from users' initial question words to queries. Different types of aids can be helpful at the steps where users must make choices.
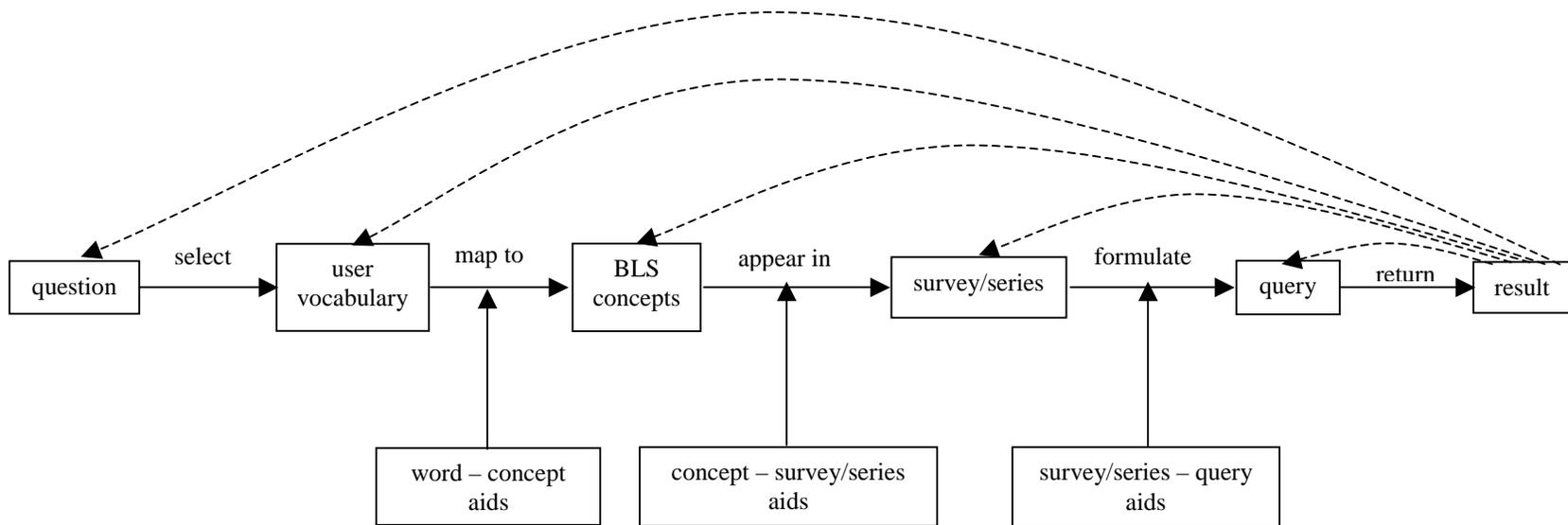
Figure 4. Feedback loops showing modifications that may be made depending on user's satisfaction with result. Examples of modifications include:

> **Query formulation – change choice of values**
> **Survey/series – use a different survey/series**
> **BLS concept – choose a related concept, e.g., industry sector rather than specific type of industry**
> **User vocabulary – use a synonym of the original word, e.g., *medical care* rather than *health care***
> **Question – simplify the question, e.g., drop *age* as a limiting concept.**

Figure 4 provides a broader view of the user's information seeking, which incorporates feedback based on the results of the user's query. The initial path, (shown with solid arrows), starts with the user's question and ends at the result returned from the database. If the user is either satisfied with the end result, or decides not to pursue a better result, this is the only path taken, and the information seeking session terminates. Otherwise, the user attempts to modify the query by changing one or more of the decisions he/she made along the way. Paths from the result back to the various decision points are shown with dotted arrows. After making a new decision(s), the user proceeds along the original path to obtain the result.

Users with knowledge about the domain, data, and/or searching techniques may base their modification(s) on the initial query result combined with their knowledge. Less expert users may make modifications in a more-or-less systematic way, making a change and evaluating the result to see if it is "closer" to the answer they want. For example, a user may first select a different value for a variable, then may try querying a different series, then may modify their initial question, and so on. Finally, a user may have no information on which to base their change(s), and may make new selections essentially at random. Either of these latter two "strategies" calls for persistence on the part of the user; if no progress is seen, he/she will become discouraged and abandon the query. The goal of user aids is to provide users with guidance tools, giving them a basis for their decisions, rather than forcing them to depend on blind chance.

This model claims that the three crucial points at which users may need additional information in order to build a successful query are at the translation or transformation points. To that end, it proposes three types of information aids: **word – concept aids**, **concept – survey/series aids**, and **survey/series – query aids**. Aids can range from simple definitions or examples to automatic parsers or focused browsing interfaces. Buckland et al. (1999) propose several tools that could be considered aids in this model. In the following discussion, examples of aids have been provided, but it is important to note that these are not exhaustive by any means, nor are they necessarily shown in the best place for the user, which should be determined through user studies. All aids should probably be available to some extent at all points in the process; the issue is which ones should be most immediately available, or even actively offered to the user. Finally, construction or adaptation of most of these aids will require more expertise in the BLS domain than I have, therefore the examples shown here are for demonstration only, and should not be assumed to be complete or error-free.

5. Components of the Model

In this section, the components of the model (the boxes shown in Figure 3) are described in more detail. Extended examples of each component, including examples of different types of aids, may be found in Appendix A. The extended examples are based on four BLS concepts: *industry*, *establishment size*, *occupation*, and *blue-collar/white-collar occupations*. Two of these are what could be considered "core" concepts in the BLS

domain (*industry* and *occupation*), the other two are related to these. Portions of the industry example (indented, in smaller print) are included in this section for illustration.

Appendix B contains tables that show the surveys/series in which the *industry* and *occupation* concepts occur, as well as the variables and values in which they are expressed. Appendix C additionally shows end user words from the LSC, and notes on the types of clarifications or aids that might be helpful. These latter two should both be considered as the raw material from which the examples were constructed.

A. User vocabulary. Sample lists of the vocabulary users have used or might use in their initial questions have been drawn from the LSC (Haas, 2000) and analysis of user queries (Liddy & Liddy, 2001). A more theoretical presentation on the relationship between end user vocabulary and BLS terminology may be found in Haas & Hert (2001).

> General words for *industry*: business, commerce, company, conglomerate, corporation, corporate entity, employer, enterprise, establishment, industry, organization
> Specific industry names, such as: airlines, beverage suppliers, coal mining, commercial banking, nursing homes, pharmaceutical manufacturers, public schools, restaurants, shoe stores, textbook publishers. Note the varying levels of specificity.

B. BLS concepts. BLS concepts are similarly drawn from the LSC, although internal BLS documents should also be mined. Concepts could be identified directly from user words, or a candidate list for the user to browse may be more helpful. "See also" concepts, which are related to the primary concept, have also been suggested in the extended examples. Showing related concepts may help the user refine his/her query in light of the BLS context and content – what is actually available, how it is organized, and how concepts can be modified or combined. The concept component is crucial because it acts as the bridge between the user's initial question and the information that the BLS can provide. This is one of the points at which the user may need to negotiate his/her question to better correspond with BLS information, or may decide that the BLS is not the right place to find the answer.

> Industry
> See also the concepts of : division, ownership, establishment, establishment size

C. Survey/series. This phrase is shorthand for all the BLS data and information products, but in the examples, I have concentrated on those available through Selective Access. The "translation" from the preceding concept component to this component is where the user can narrow down the list of available survey/series to those that are relevant to the concept(s) of interest. The Matrix Model of Survey/Series and Variables (Section 3) demonstrates how this could be accomplished.

> *Industry* Survey/series
> 1. Labor Force Statistics from the Current Population Survey
> 2. Nonfarm Payroll Statistics from the Current Employment Statistics (National)
> 3. Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)
> 4. Covered Employment and Wages
> 5. Mass Layoff Statistics
> 6. Producer Price Index Revision – Current Series
> 7. Producer Price Index Revision – Discontinued Series

8. Collective Bargaining Statistics – Historical (Private Sector)
9. Collective Bargaining Statistics – Historical (State and Local Government)
10. Employee Benefits Survey
11. Employment Cost Trends (Employment Cost Index)
12. Nonfatal Cases Involving Days Away From Work: Selected Characteristics
13. Census of Fatal Occupational Injuries
14. Occupational Injuries and Illnesses: Industry Data (pre 1989)
15. Occupational Injuries and Illnesses: Industry Data (1989-current)
16. Major Sector Productivity and Costs Index
17. Major Sector Multifactor Productivity Index
18. International Labor Statistics

The extended examples in Appendix A are presented as though a single concept was involved, but of course many questions contain more than one. In this case, the list of relevant survey/series could at first be assumed to be the intersection of those relevant to each concept individually. This may not always be the best strategy, however. If the intersection is empty, indicating that the concepts do not all appear in the same survey/series, then the user needs to be able to "back out" of the selection, and either discard one or more of the concepts or seek information somewhere other than the BLS.

The choice of survey/series is a complex one, involving the kind of information the user is seeking, the level of detail at which the query should be expressed, limitations in the survey on which a series is based, and a myriad of other factors. Although in some cases a user interface may be able to "suggest" one or more as being the best (or only) choice, the decision is ultimately made by the user, who may take into account any compromises between the initial question and the final query.

D. Word – concept aids. The purpose of the word – concept aids is to help users select the BLS concepts that best correspond to the words and concepts in their initial question. The following list suggests some types of aids that may be helpful.

- General definition of concept, preferably drawn from BLS documentation, written in non-technical terms.

- Ambiguity resolution. If the user's word(s) could correspond to more than one concept, allow the user to select the correct concept. This could mean pointing out an ambiguity that the user did not notice, especially if it occurs because of the nature of BLS terminology. For example, if a user asks about *plumbing*, he/she could be referring to an occupation or an industry. The BLS is more precise in maintaining the distinction between *plumbing* (industry) and *plumber* (occupation). Another example is the word *sector*, which may refer to a group of industries, or to the distinction between public and private ownership of an establishment (e.g., *manufacturing sector* or *private sector*).

- Synonyms. Other terms that may be used for the concept name within the BLS. The BLS terminology is not "perfect", with a one-to-one relationship between words and meanings. Different terms are occasionally used for the same concept in different resources, and the same term is sometimes used for different concepts.

Industry synonyms

Sector, product group (in some cases)

Also, see the variable names used within the series, and those used for the see-also concepts listed in B.

- Examples. Examples of queries that can be asked involving that concept. This could be similar to an "Ask Jeeves" kind of resource, in which the examples illustrate the kind of information available. A Frequently Asked Question list (FAQ) would be another way of presenting examples.

    Industry example queries

    How many lost workday due to injury were there in the construction industry last year?

    What were the average hourly earnings of workers in goods-producing industries for the last 5 years?

    How many people are employed as insurance agents?

    Other kinds of examples could illustrate the range of possible values, e.g., demonstrating the structure of a classification.

    Examples of 2-, 3-, and 4-digit SIC codes and industries, to give users the idea of varying levels of specificity.

    a. Division D: Manufacturing

    Major Group 25: Furniture And Fixtures

    Industry Group 251: Household Furniture

    2511 Wood Household Furniture, Except Upholstered

    b. Division G: Retail Trade

    Major Group 56: Apparel And Accessory Stores

    Industry Group 566: Shoe Stores

    5661 Shoe Stores

    c. Division I: Services

    Major Group 80: Health Services

    Industry Group 805: Nursing And Personal Care Facilities

    8052 Intermediate Care Facilities

- Parsing. Phrases or questions could be automatically parsed into concepts, or even queries, in cases where there is little room for ambiguity.

    Industry parsing examples

    For general industry words listed in A, show user descriptions of surveys/series listed in E2. Or add an intermediate step, first allowing user to choose industry concept or one of the see-also concepts listed in B. (Use similar treatment for *occupation*.)

    Always have options of browsing SIC available to user. It is not feasible to maintain a lexicon of all possible words naming specific industries or occupations.

- Thesaurus browsing. A thesaurus can present additional concepts that stand in a variety of relationships to the initial one. A domain-specific thesaurus could serve as a means by which users could increase their familiarity with the

information the BLS has available, as well as learning about concepts specifically related to those in the question.

- Links to references. Reference sources such as the SIC or SOC classifications could help users understand how the "meaning" of terms such as *industry* and *occupation* is structured in the BLS domain.

    > *Industry* links to authoritative classification
    > Link to Standard Industry Classification, along with brief description of its structure and use. http://www.osha.gov/cgi-bin/sic/sicser5
    > Tree diagram of the major groupings (i.e., divisions) of SIC.

    Other types of references could highlight the relationship between concepts found in familiar reference sources and BLS definitions.

    > *Industry* links to familiar reference sources
    > Create crosswalks linking familiar, general purpose categorizations of industries with SIC/NAICS categories. Two candidates are given below.
    > The Yahoo B2B directory
    > (http://dir.yahoo.com/Business_and_Economy/Business_to_Business/), which includes categories such as aerospace, architecture, automotive, chemicals and allied products, electronics, food and beverage, furniture, health care.
    > The Yahoo Consumer Yellow Pages (http://yp.yahoo.com/), whose top-level categories are automotive, community, computers and internet, education and instruction, entertainment and arts, food and dining, health and medicine, home and garden, legal and financial, other shopping and services, personal care, real estate, recreation and sports, travel and transportation.

- Scope notes. These notes would explicitly describe meaning or limitations of coverage of a term or concept in the BLS domain.

- Links to BLS documentation. Documentation such as the *Handbook of Methods* (http://www.bls.gov/opub/hom/homhome.htm) is probably not of interest to the casual BLS user, but can provide a deeper understanding of terms and concepts in the BLS domain. These links would be most useful if they could be context-sensitive, pointing to specific chapters or sections depending on the concept under consideration. However, even pointing out the existence of a source, and allowing the user to browse or query for him/herself is useful. Links should be annotated to indicate the type of resource or information that will be reached.

E. Concept – survey/series aids. The purpose of these aids is to help the users identify survey/series that are relevant to the concept(s) in their questions, that is, those in which the concept appears in some form. The following list suggests some types of aids that may be helpful. Examples from Appendix A follow the list.

- Selection of survey/series that contain the concept. This can be done automatically based on the Matrix discussed in Section 3. For some concepts, such as *industry*, there will still be many from which to choose; for others, such as *establishment size*, there will only be one or two remaining.

- Summary of information available in relevant survey/series. The summary should give the user a general idea of the kinds of questions that the survey can answer. To the greatest extent possible, the summary should be written in non-technical language, so as to be meaningful to non-experts. The overviews found in the *Series ID Formats* (http://stats.bls.gov/hlpforma.htm) provide a good starting point.

- List of variables available in the relevant surveys. The list would help users understand the context in which the concept(s) of interest are presented. Brief definitions would be necessary for those whose names are not self-explanatory (such as datatype). Variables that correspond to the concept(s) of interest should be highlighted.

- Scope notes. These scope notes would describe restrictions or limitations on the concepts are they are represented in each relevant survey/series. For example, the general BLS concept of *industry* is quite inclusive, but in some surveys/series, agriculture is omitted.

- Link to documentation. Links here could be to survey/series-specific overviews, documentation, or technical reports. Links should be annotated to indicate the type of resource or information that will be reached.

- Links to other sources of related information. Frequently, users search for information on the BLS website that is not in its purview. For example, the BLS does collect data on workplace injuries, but does not enforce workplace safety regulations. A concept-survey/series aid associated with the injury and fatality series could refer users to the appropriate federal agencies. Naturally, the BLS cannot provide this kind of service for all the information it does not have, but there are many frequently occurring confusions that could be helped in this way.

Four examples from the eighteen given in *Appendix A, Example I, Concept – survey/series aids* are shown below. Each example includes:
- A 1-2 sentence summary of the information available in the survey/series, with link(s) to additional documentation, e.g., the relevant survey overview, or the *Handbook of Methods*.
- Definition and/or scope of *industry* concept as used in the survey/series, especially noting any restrictions in coverage.
- List of all variables available in the survey/series, with definition where necessary. (E.g., from Series ID Formats, http://stats.bls.gov/hlpforma.htm.)
- Highlighted variables that are related to industry. (These variables are underlined below.)
Also listed are any possible sources of related data, in case these resources can't satisfy the user's information need.

> ***Labor Force Statistics from the Current Population Survey***
> **Summary.** This series presents information about the size and makeup of the U.S. labor force, which can be broken out by several demographic variables.
> **Definition and Scope**. *Industry* covers all industries, including private households, at the division, division combination, down to the 3-digit SIC level.

Variables. Seasonal Adjustment, Age, Sex, Race, Ethnicity, Occupation,
class of worker - (sector in which individual works),
status- (portion of labor force included),
industry - (division or industry in which individual works).
Other Sources. Current Population Survey, Bureau of the Census

Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)
Summary. The Current Employment Statistics program is a monthly survey conducted by State employment security agencies in cooperation with the Bureau of Labor Statistics. The survey provides employment, hours and earnings estimates based on payroll records of business establishments. (http://stats.bls.gov/cesprog.htm)
Definition and Scope. Large number of values, including division, division combinations, and 2-4-digit SIC codes. Excludes agriculture ("nonfarm" only).
Variables. Seasonal adjustment, state, area, data type (indicates the unit of measurement, such as number of employees, earnings, hours worked),
detail - (level of specificity for presenting industry values,
industry - (SIC-based industries).
Other Sources.

Mass Layoff Statistics
Summary. Information is obtained on the total number of persons separated; the reasons for separation; worksite closures; recall expectations; and socioeconomic characteristics on UI claimants – such as gender, age, race, and residency.
Definition and Scope. Industry-related values are large-scale groupings of industry types. Some reasons for layoff are marginally industry-related, such as "model changeover" or "seasonal work".
Variables. Seasonal adjustment, period (monthly or quarterly), SRD (geographical area of state, region, division), datatype (layoff events, number of separations, total initial claimants, rate),
IRC - Industry, Reason and Characteristics. Contains values for industry, reason for layoff (e.g., bankruptcy), and demographic characteristics of claimants.
Other Sources.

Nonfatal Cases Involving Days Away From Work: Selected Characteristics
Summary. Since 1972, the survey has reported annually on the number of workplace injuries and illnesses in private industry and the frequency of those incidents.
Definition and Scope. Private industry only, available at division and 3-digit SIC levels.
Variables. Seasonal adjustment, data type (rate of injury/illness, number of cases, or median workdays lost),
case type - (how data is broken out, including industry division by source of illness/injury, event, etc., or injury/illness by detailed industry),
category - (identifies the broad industry division, injury and illness classification, or demographic characteristic),
industry - (industry divisions plus 2- and 3-digit SIC).
Other Sources.

F. Survey/series – query aids. The purpose of these aids is to help users formulate the query once they have chosen a survey/series. These could include query construction tools similar to Public Data Query, or even tools that automatically construct the query,

or part of it, based on the decisions the user has already made. Examples from Appendix A follow this list.

- Lists of relevant variables and their values. It would be helpful for users to be able to browse an organized, nicely displayed list of values for a given variable. Values that could be confusing, such as "a residual cell" (one of the values for detail in *Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)*) could be linked to definitions. In cases where there are hundreds of values, such as the industry variable in the same series, it might be helpful to also provide a link to the classification on which it is based (in this case, the SIC). As a final note, in cases where the variable in the original survey differs from the one(s) presented in the query interface, the difference should be noted. For example, the data dictionary for *Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)* lists a single variable called industry, but the Selective Access tool divides it into detail and industry, perhaps due to the number of possible values.

- Constraints or interactions between variable-value choices. As mentioned earlier, frequently the selection of a value for one variable constrains the list of possible values for another. Explaining these constraints helps the user find the correct "balance" between choices, and also prevents the possibly unnerving "loss" of possible values behind the scenes. These explanations could also help prevent null query results, when no data is returned to the user.

- Notes on query results. These notes would warn the user of unusual situations in the data that might result from the query. Examples include the presence of preliminary data, sparse data, breaks or changes in the administration of the survey, etc. Currently this information is frequently found in the footnotes to a table or a series.

- Links to references or documentation. As before.

- Query construction interfaces. These could include the current Selective Access and Public Data Query, as well as others yet to be developed. If there is a choice of tools, annotation should help guide the user to the best one for his/her situation.

Four examples from the eighteen given in *Appendix A, Example I, Survey/series – query aids* corresponding to those given in the preceding section, are shown below. Each example includes:
- Lists of values for industry-related variables. Note: literal values are in small font, descriptions of values are in normal size.
- Description of constraints or interactions between industry-related variable-value choices.
- Notes on possible issues with results, e.g., breaks in collection, preliminary data, changes in definition.

Labor Force Statistics from the Current Population Survey
Industry-related variables and values.

class of worker N/A, Wage and Salary Workers, Private Wage and Salary Workers, Government Wage and Salary Workers, Self-employed Workers, Unpaid Family Workers, Self-employed And Unpaid Family Workers (8 and 9)

status Civilian Labor Force, Total Labor Force (Includes Total Armed Forces), Full-time Labor Force, Part-time Labor Force, Armed Forces, Not in Labor Force, N/a, Civilian Labor Force – ½ of Part Time Labor Force, Code 7 + Discouraged Workers (den. Of U-7 = Den. O, Labor Force (Includes Resident Armed Forces.)

industry

Private households (meaning private households in nonagricultural private wage and salary)
Nonagriculture goods producing industries (mining, construction, and manufacturing)
Service producing industries (nonag, code 040, except mining, construction and manufacturing)
Nonagricultural industries (meaning: mining; constru; manufacturing; trans and public utilities; trade; fin, insurance, and real estate; sv industries including forestry and fisheries; and public admin)
Construction
Manufacturing
Manufacturing, durables
Manufacturing, nondurables
Transportation, communication, and other public utilities (transportation and public utilities)
Trade (wholesale and retail)
Finance and service industries (meaning sv ind, 406, and tp26 ind codes 707-709)
Service industries (includes forestry and fisheries and private households)
Finance, insurance, and real estate
Service industries except private households (miscellaneous services on tabs, E.g., 55)
All nonagricultural industries except private households (this code has also been used in conjunction with 5:1 and 19:2 to identify other nonagriculture private wage and salary than private households, e.g., tabs 15)
Also, long list of 3-digit SIC level industries.

Constraints or Interactions. To get choices of detailed SICs, status must equal civilian labor force, other variables must equal null.

Notes.

Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)
Industry-related variables and values.

detail Total non-agricultural, 1-digit industry (Industry division), 2-digit industry
3-digit industry, A combination of industry divisions, A combination of industries, A residual cell

industry. All divisions, 2- and 3-digit SIC codes, plus

goods producing
service producing
construction and mining
services and mining (has asterisk)
Durable goods
Lumber & wood products including furniture
Prim metal inds & fabr. Metal prods
33-36
mach & electronic & electric equip.
nondurable goods
textile mill products & apparel
broadwoven and narrow fabric mills
men's & boys' suits, coats, & furngs.
Paper & printing & publishing
Pulp, paper and paperboard mills
261, 262, 263, 266
chemicals & petroleum & coal products
communications and public utilities
wholesale trade
retail trade
*R
Other durable goods
Other nondurable goods
Other federal government
Other state and local government

Constraints or Interactions. Value chosen for detail controls values displayed for industry. If seasonally adjusted, fewer values available.

Notes. Need explanation of detail value "a residual cell", and the industry value "*R".

Mass Layoff Statistics
Industry-related variables and values.

<u>IRC</u> Total all industries, Total all industry, private, Agriculture, Nonagriculture, Manufacturing, Durable goods
(also list of specific durable goods industries, including "miscellaneous manufacturing industries")

Nondurable goods
(also list of specific nondurable goods industries)

Non manufacturing
(also list of divisions)

Not identified, Government, Federal, State, Local

Total, all reasons, Automation, bankruptcy, business ownership change, contract cancellation, contract completion, domestic relocation, energy-related, environment-related, financial difficulty, import competition, labor dispute, material shortage, model changeover, natural disaster, overseas relocation, plant or machine repair, product line discontinued, reorganization within company, seasonal work, slack work, vacation period, weather-related, other reasons, reason not reported

Characteristics = black, Hispanic, women
Constraints or Interactions.
Notes. IRC contains many kinds of information, including ownership as well as industry type. Also includes non-industry-related choices.

Nonfatal Cases Involving Days Away From Work: Selected Characteristics
Industry-related variables and values.

<u>case type</u>
Industry division or selected characteristic by detailed nature of condition
Industry division or selected characteristic by detailed part of body affected
Industry division or selected characteristic by detailed source of injury/illness
Industry division or selected characteristic by detailed event or exposure
Industry division or selected characteristic by detailed occupation
Selected injury/illness characteristic by detailed industry
Industry division or length of absence by age group
Industry division or length of absence by gender
Industry division or length of absence by race
Industry division or length of absence by length of service

<u>category</u>
Total private industry
Industry division – agriculture
Industry division – mining
Industry division – construction
Industry division – manufacturing
Industry division – transportation and public utilities
Industry division – wholesale trade
Industry division -- retail trade
Industry division – finance, insurance, and real estate
Industry division – services

Also includes values for age groups, days away from work, event or exposure, gender, length of service, nature of injury, occupation, part of body affected, race, source of injury.

<u>industry</u> Agriculture, forestry, and fishing, Mining, Construction, Manufacturing, Transportation, communications, electric, gas, and sanitary services, Wholesale trade, Retail trade, Finance, insurance, and real estate, Services

Plus 2- and 3-digit SIC codes.
Constraints or Interactions. Detailed industry values available only if case type = detailed industry.
Notes. The order in which occupations are listed seems to be based on codes, which results in some very odd sequences. Can this be displayed as a drill-down list?

6. Summary

The work described in this report has explored ways of improving end user access to BLS information, with two main emphases.
1. How to bridge the gap between users' concepts and vocabulary and the concepts and terms used in the BLS domain.
2. How to provide information to the users at each stage in the search process that will inform them about the information available, and enable them to make good decisions in formulating queries. This process includes identifying the BLS concepts that correspond to the concepts in their queries, selecting survey/series that are relevant to these concepts, and selecting variable-value pairs.

The two models that were created during the course of this work illustrate where the crucial decision points are in the user's information-seeking process, and demonstrate some means by which useful information could be provided.

Many of the information resources required to implement these models already exist within the agency. The would need to be identified and located, and incorporated into a coherent tool set under a good user interface. Others may need to be created; an incremental development process is most realistic for these, with priority given to those addressing the most common user problems.

Finally, it should be noted that any of the aids suggested here would require ongoing maintenance. For many, such as the Matrix model, changes would mirror those made in other BLS information resources (e.g., changing from SIC to NAICS), and would be relatively stable over time. Others would require more constant effort. For example, the LSC requires tracking changes in users' vocabulary and query concepts by sampling actual queries and environmental scanning. The long-term benefit of continuing to improve service to the public by providing them with better answers to their questions, as well as opportunities to learn about BLS data and how to use it, should compensate for these efforts.

The following products were prepared in the course of this work:
- Harmonization of the LABSTAT Crosswalk semantic groupings (Haas, 2000) and the Liddy & Liddy (2001) grammar categories (Section 2).
- Matrix Model of Survey/Series and Variables, first presented in the May 14, 2001 progress report, along with ideas for its user in supporting user queries (Section 3).
- Model of the Progression from Question to Query Formulation, which suggests information aids that could be useful at each stage (Sections 4 and 5).
- Four extended examples of how the Progression Model can help organize and present helpful information to end users (Appendix A).
- Background information showing the LABSTAT surveys/series and variables in which the *industry* and *occupation* concepts occur (Appendix B).
- Background information mapping end user words and phrases describing *industry* and *occupation* to BLS surveys/series and variables (Appendix C).

Additional papers and presentations include:

> Haas, S. W. & Hert, C. A. (2001). Finding information at the U. S. Bureau of Labor Statistics: Overcoming the barriers of scope, concept, and language mismatch. Submitted for publication.

> Haas, S. W. & Hert, C. A. (2001). Sharing the blame: Understanding retrieval performance in terms of language systems and information availability. In preparation.

> Haas, S. W. (2001). Maping user concepts and words to BLS concepts and terms, or I don't care what you call it, just give me the answer! Presentation as member of panel organized by Deborah Klein, *Incorporating User Input into the Design of a Website: Research and Implementation (SIG USE)*, Annual Meeting of the American Society for Information Science and Technology, November 5, 2001.

7. Bibliography

Buckland, M., Chen, A., Chen, H., Kim, Y., Lam, B., Larson, R., Norgard, B., Purat, J. & Gey, F. (1999). Mapping entry vocabulary to unfamiliar metadata vocabularies. *D-Lib Magazine*, 5, 1, http://www.dlib.org/dlib/january99/buckland/01buckland.html.

Haas, S. W. (2000). A Terminology Crosswalk for LABSTAT: Mapping General Language Words and Phrases to BLS Terms. Final report for the Bureau of Labor Statistics contract, "Terminology Crosswalks for LABSTAT". http://ils.unc.edu/~stephani/bls/fin-rept.pdf.

Haas, S. W. & Hert, C. A. (2001). Finding information at the U. S. Bureau of Labor Statistics: Overcoming the barriers of scope, concept, and language mismatch. Submitted for publication.

Liddy, E. & Liddy, J. (2001). An NLP Approach for Improving Access to Statistical Information for the Masses. Paper to be presented to the *FCSM 2001 Research Conference* (Nov, 2001). http://istweb.syr.edu/~tables

Marchionini, G. (2000). Interfaces to support customized views and manipulation of statistical data. The second international conference on establishment surveys (Buffalo, NY, June 17-21, 2000). Alexandria, VA: American Statistical Association. 953-959.

Marchionini, G., Brunk, B., Komlodi, A., Conrad, F., & Bosley, J. (2000). Look Before You Click: A Relation Browser for Federal Statistics Websites. Proceedings of the Annual Meeting of the American Society for Information Science (Chicago, Nov. 12-16, 2000), 392-402.

Appendix A

Extended Examples of Based on the Progression Model

I. The Industry Concept
II. The Establishment Size Concept
III. The Occupation Concept
IV. The White Collar/Blue Collar Concept

Each example includes the following components:
       A. User Vocabulary Examples
       B. BLS Concepts
       C. Surveys/Series
       D. Word-Concept Aids
       E. Concept-Survey/Series Aids
       F. Survey/Series-Query Aids

Note: These examples are provided for illustrative purposes only, and should not be assumed to be complete or error-free. Construction of most of the included aids will require expertise in the BLS domain.

Extended Example I: The Industry Concept

This concept encompasses both the meaning of *industry* as a whole, (e.g., "give me unemployment figures for January 2001 by industry") and specific industry types, (e.g., "How many people work in commercial real estate"). This concept was chosen in part because of the number of survey/series in which it appears. Related concepts are given in section B.

A. User Vocabulary Examples.
business, commerce, company, conglomerate, corporation, corporate entity, employer, enterprise, establishment, industry, organization
Specific industry names, such as: airlines, beverage suppliers, coal mining, commercial banking, nursing homes, pharmaceutical manufacturers, public schools, restaurants, shoe stores, textbook publishers. Note the varying levels of specificity.

B. BLS Concept
Industry
See also the concepts of: division, ownership, establishment, establishment size

C. Surveys/Series
1. Labor Force Statistics from the Current Population Survey
2. Nonfarm Payroll Statistics from the Current Employment Statistics (National)
3. Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)
4. Covered Employment and Wages
5. Mass Layoff Statistics
6. Producer Price Index Revision – Current Series
7. Producer Price Index Revision – Discontinued Series
8. Collective Bargaining Statistics – Historical (Private Sector)
9. Collective Bargaining Statistics – Historical (State and Local Government)
10. Employee Benefits Survey
11. Employment Cost Trends (Employment Cost Index)
12. Nonfatal Cases Involving Days Away From Work: Selected Characteristics
13. Census of Fatal Occupational Injuries
14. Occupational Injuries and Illnesses: Industry Data (pre 1989)
15. Occupational Injuries and Illnesses: Industry Data (1989-current)
16. Major Sector Productivity and Costs Index
17. Major Sector Multifactor Productivity Index
18. International Labor Statistics

D. Word-Concept Aids
1. General definition of *industry* concept as used in the BLS. (Drawn from BLS document.)

2. Distinction between industry and occupation. An industry is a public or private organization or institution that … An occupation is the job or career that an individual follows, usually in return for compensation of some sort. In general language, the same word is often used for both, e.g., *farming*, or *plumbing*. In the BLS data, however, there is a distinction between the industry of *agriculture* and the occupation of *farming*, and a similar distinction between *plumbing* and *plumbers*. (Also use this clarification of the *occupation* concept.)

3. Synonyms and related terms within the BLS. See, for example, the variable names used within the series, and those used for the see-also concepts listed in B.
sector, product group (in some cases)

4. Link to Standard Industry Classification, along with brief description of its structure and use.
http://www.osha.gov/cgi-bin/sic/sicser5

5. Tree diagram of the major groupings (i.e., divisions) of SIC. (Note: this would also be useful for the *division* concept. Does it already exist somewhere?)

6. Examples of 2-, 3-, and 4-digit SIC codes and industries, to give users the idea of varying levels of specificity.
>   a. Division D: Manufacturing
>>   Major Group 25: Furniture And Fixtures
>>>   Industry Group 251: Household Furniture
>>>>   2511 Wood Household Furniture, Except Upholstered
>   b. Division G: Retail Trade
>>   Major Group 56: Apparel And Accessory Stores
>>>   Industry Group 566: Shoe Stores
>>>>   5661 Shoe Stores
>   c. Division I: Services
>>   Major Group 80: Health Services
>>>   Industry Group 805: Nursing And Personal Care Facilities
>>>>   8052 Intermediate Care Facilities

7. Create crosswalks linking familiar, general purpose categorizations of industries with SIC/NAICS categories. Two candidates are given below.
>   The Yahoo B2B directory
>   (http://dir.yahoo.com/Business_and_Economy/Business_to_Business/), which includes categories such as aerospace, architecture, automotive, chemicals and allied products, electronics, food and beverage, furniture, health care.
>   The Yahoo Consumer Yellow Pages (http://yp.yahoo.com/), whose top-level categories are automotive, community, computers and internet, education and instruction, entertainment and arts, food and dining, health and medicine, home and garden, legal and financial, other shopping and services, personal care, real estate, recreation and sports, travel and transportation.

8. Parsing possibilities.
- For general industry words listed in A, show user descriptions of surveys/series listed in E2. Or add an intermediate step, first allowing user to choose industry concept or one of the see-also concepts listed in B. (Use similar treatment for *occupation*.)
- Always have options of browsing SIC available to user. It is not feasible to maintain a lexicon of all possible words naming specific industries or occupations.

9. Example questions, either illustrating frequently asked questions (FAQs), or to demonstrate how variables/values can be combined.
>   How many lost workday due to injury were there in the construction industry last year?
>   What were the average hourly earnings of workers in goods-producing industries for the last 5 years?
>   How many people are employed as insurance agents?

Appendix A   3

E. Concept-Survey/Series Aids
For each relevant survey/series (see C):
- A 1-2 sentence summary of the information available in the survey/series, with link(s) to additional documentation, e.g., the relevant survey overview, or the *Handbook of Methods*.
- Definition and/or scope of *industry* concept as used in the survey/series, especially noting any restrictions in coverage.
- List of all variables available in the survey/series, with definition where necessary. (E.g., from Series ID Formats, http://stats.bls.gov/hlpforma.htm.)
- Highlight those variables related to industry. (These variables are underlined below.)

Also list any possible sources of related data, in case these resources can't satisfy the user's information need.

1. Labor Force Statistics from the Current Population Survey
Summary. This series presents information about the size and makeup of the U.S. labor force, which can be broken out by several demographic variables.
Definition and Scope. *Industry* covers all industries, including private households, at the division, division combination, down to the 3-digit SIC level.
Variables. Seasonal Adjustment, Age, Sex, Race, Ethnicity, Occupation,
> class of worker - (sector in which individual works),
> status- (portion of labor force included),
> industry - (division or industry in which individual works).

Other Sources. Current Population Survey, Bureau of the Census

2. Nonfarm Payroll Statistics from the Current Employment Statistics (National Employment, Hours, and Earnings)
Summary. The Current Employment Statistics program is a monthly survey conducted by State employment security agencies in cooperation with the Bureau of Labor Statistics. The survey provides employment, hours and earnings estimates based on payroll records of business establishments. (http://stats.bls.gov/cesprog.htm)
Definition and Scope. Large number of values, including division, division combinations, and 2-4-digit SIC codes. Excludes agriculture ("nonfarm" only).
Variables. Seasonal adjustment, data type (indicates the unit of measurement, such as number of employees, earnings, hours worked, diffusion indexes),
> industry- (SIC-based industries).

Other Sources.

3. Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)
Summary. The Current Employment Statistics program is a monthly survey conducted by State employment security agencies in cooperation with the Bureau of Labor Statistics. The survey provides employment, hours and earnings estimates based on payroll records of business establishments. (http://stats.bls.gov/cesprog.htm)
Definition and Scope. Large number of values, including division, division combinations, and 2-4-digit SIC codes. Excludes agriculture ("nonfarm" only).
Variables. Seasonal adjustment, state, area, data type (indicates the unit of measurement, such as number of employees, earnings, hours worked),
> detail - (level of specificity for presenting industry values,
> industry - (SIC-based industries).

Other Sources.

4. Covered Employment and Wages

Summary. Includes data on the number of establishments, monthly employment, and quarterly wages, by industry, at the 4-digit Standard Industrial Classification (SIC) level, by county, by ownership sector, for the entire United States. These data are aggregated to annual levels, to higher industry levels (3-digit, 2-digit, industry division and totals), and to higher geographic levels (national, state, and Metropolitan Statistical Area (MSA)).

Definition and Scope. Large number of values for industry, including "nonclassifiable establishments" Can be coordinated with ownership for government/private distinction.

Variables. Seasonal adjustment, area, datatype,

> size - Size of establishment. (Currently no choices are available, but that is expected to change),
> ownership - Distinguishes between various governmental levels and private,
> industry - (SIC values to 4 digits).

Other Sources.

5. Mass Layoff Statistics

Summary. Information is obtained on the total number of persons separated; the reasons for separation; worksite closures; recall expectations; and socioeconomic characteristics on UI claimants – such as gender, age, race, and residency.

Definition and Scope. Industry-related values are large-scale groupings of industry types. Some reasons for layoff are marginally industry-related, such as "model changeover" or "seasonal work".

Variables. Seasonal adjustment, period (monthly or quarterly), SRD (geographical area of state, region, division), datatype (layoff events, number of separations, total initial claimants, rate),

> IRC - Industry, Reason and Characteristics. Contains values for industry, reason for layoff (e.g., bankruptcy), and demographic characteristics of claimants.

Other Sources.

6. Producer Price Index Revision – Current Series

Summary. The Producer Price Index (PPI) is a family of indexes that measures the average change over time in selling prices received by domestic producers of goods and services. PPIs measure price change from the perspective of the seller.

Definition and Scope. Extensive list of industry values at varying levels of SIC codes. Also industry groupings associated with stage of processing. A product is defined within a specific industry.

Variables. Seasonal adjustment,

> industry - 2-4-digit SIC codes, plus stage of processing values,
> product - detailed product codes are the numerical extensions to the industry codes at the 5-6-digit level. Product groupings across SIC codes are also available.

Other Sources.

7. Producer Price Index Revision – Discontinued Series

NOTE: Same as PPI Current, except slightly shorter lists of values for industry and product.

Summary.
Definition and Scope.
Variables.
Other Sources.

8. Collective Bargaining Statistics – Historical (Private Sector)
Summary. This series summarizes wage-rate changes in major collective bargaining agreements for production and related workers in manufacturing and nonsupervisory workers in non-manufacturing.
Discontinued after 1995.
Definition and Scope. Industries covered by collective bargaining agreements of 1,000 or more workers.
Variables. Measure (e.g., wage rates for 1st year of contract, compensation costs over life of contract), alteration (always changes?), unit (always mean?),
      <u>industry</u> - just a few broad industry combinations.
Other Sources.

9. Collective Bargaining Statistics – Historical (State and Local Government)
Summary. This series summarizes wage-rate changes in major collective bargaining units in state and local governments where: (1) A labor organization is recognized as the bargaining agent for a group of workers; (2) the settlements are embodied in signed, mutually binding contracts; and (3) wages are determined by collective bargaining.
Discontinued after 1995.
Definition and Scope. State and local governments covered by collective bargaining agreements of 1,000 or more workers.
Variables. Measure (e.g., wage rates for 1st year of contract, compensation costs over life of contract), alteration (always changes?), unit (always mean?),
      <u>sector</u> - state and/or local governments.
Other Sources.

10. Employee Benefits Survey
Summary. The Employee Benefits Survey (EBS) of the Bureau of Labor Statistics (BLS) covers the incidence and characteristics of employee benefit plans.
Definition and Scope. Covers private and government, but no specific industry type breakdowns.
Variables. Seasonal adjustment, title (category of benefit provision),
      <u>type</u> - (size and type of establishment).
Other Sources.

11. Employment Cost Trends (Employment Cost Index)
Summary. The ECI measures the change over time in the cost of labor. The cost of labor includes wages and salaries and employer costs for employee benefits.
Definition and Scope. Industry is represented by combined divisions, some subdivisions, and some very specific (4-digit SIC) values.
Variables. Seasonal adjustment, compensation (wage, benefits, or both), periodicity (quarterly or annual average),
      <u>group</u> - (industry, occupation, region, union status of compensation – some in
      combination),
      <u>ownership</u> - sector of industry owner.
Other Sources.

12. Nonfatal Cases Involving Days Away From Work: Selected Characteristics
Summary. Since 1972, the survey has reported annually on the number of workplace injuries and illnesses in private industry and the frequency of those incidents.
Definition and Scope. Private industry only, available at division and 3-digit SIC levels.
Variables. Seasonal adjustment, data type (rate of injury/illness, number of cases, or median workdays lost),

case type - (how data is broken out, including industry division by source of illness/injury, event, etc., or injury/illness by detailed industry),

category - (identifies the broad industry division, injury and illness classification, or demographic characteristic),

industry - (industry divisions plus 2- and 3-digit SIC).

Other Sources.

13. Census of Fatal Occupational Injuries

Summary. Provides information on 28 separate data elements including information on the worker, the fatal incident, and the machinery or equipment involved.

Definition and Scope. Industry covers both private and government industry.

Variables. Data type (number of fatal injuries),

case type - (how data is broken out, including industry and sector, event, occupation, source of injury),

category - (identifies the broad industry division, injury and illness classification, or demographic characteristic),

industry - (industry divisions plus 2- and 3-digit industries).

Other Sources.

14. Occupational Injuries and Illnesses: Industry Data (pre 1989)

Summary. The number of workplace injuries and illnesses in private industry and the frequency of those incidents.

Definition and Scope. Private industry only.

Variables. Seasonal adjustment, data type (injury, illness or both), case type (lost workdays, days with restricted work activity, etc.),

division - (SIC divisions),

industry - (2-, 3-, and 4-digist SIC codes).

Other Sources.

15. Occupational Injuries and Illnesses: Industry Data (1989-current)

Summary. The number of workplace injuries and illnesses in private industry and the frequency of those incidents.

Definition and Scope. Private industry only.

Variables. Seasonal adjustment, data type (injury, illness or both), case type (lost workdays, days with restricted work activity, etc.),

division - (SIC divisions),

industry - (2-, 3-, and 4-digist SIC codes).

Other Sources.

16. Major Sector Productivity and Costs Index

Summary. Quarterly labor productivity for the major U.S. economic sectors.

Definition and Scope. A variety of sector/division combinations, private ownership only.

Variables. Seasonal adjustment, type (all persons or all employees), measure (measure of productivity e.g., employment or unit labor costs), duration (%change or index values),

sector - (SIC divisions in combination. Note: does not mean ownership).

Other Sources.

17. Major Sector Multifactor Productivity Index

Summary. Productivity and cost indexes for business, especially manufacturing.

Definition and Scope. Private ownership only, concentration on manufacturing.

Variables. Measure (type index, e.g., output/hour, cost of capital,

Other Sources.

18. International Labor Statistics
Summary. Productivity and unit labor costs; hourly compensation costs of
manufacturing production workers; the labor force, employment, and unemployment; gross
domestic product per capita and per employed person; and consumer prices and other measures.
The measures compiled relate primarily to the major industrial countries, but other countries or
areas of importance to U.S. foreign trade are included in some of the measures. Most of the series
are prepared on an annual average basis; comparative figures on unemployment and consumer
prices are prepared on a monthly and quarterly basis, as well.
Definition and Scope. Limited to manufacturing.
Variables. Seasonal adjustment, country, suffix (applies only for consumer price indexes),
group - (type of statistic or index).
Other Sources.

F. Survey/Series- Query Aids
For each relevant survey/series (see C):
- Lists of values for industry-related variables. Note: literal values are in small font, descriptions
  of values are in normal size.
- Description of constraints or interactions between industry-related variable-value choices.
- Notes on possible issues with results, e.g., breaks in collection, preliminary data, changes in
  definition.

1. Labor Force Statistics from the Current Population Survey
Industry-related variables and values.

class of worker N/A, Wage and Salary Workers, Private Wage and Salary Workers, Government Wage and Salary Workers,
Self-employed Workers, Unpaid Family Workers, Self-employed And Unpaid Family Workers (8 and 9)

status Civilian Labor Force, Total Labor Force (Includes Total Armed Forces), Full-time Labor Force, Part-time Labor Force,
Armed Forces, Not in Labor Force, N/a, Civilian Labor Force – ½ of Part Time Labor Force, Code 7 + Discouraged Workers (den. Of
U-7 = Den. O, Labor Force (Includes Resident Armed Forces.)

industry
Private households (meaning private households in nonagricultural private wage and salary)
Nonagriculture goods producing industries (mining, construction, and manufacturing)
Service producing industries (nonag, code 040, except mining, construction and manufacturing)
Nonagricultural industries (meaning: mining; constru; manufacturing; trans and public utilities; trade; fin, insurance, and real estate; sv
industries including forestry and fisheries; and public admin)
Construction
Manufacturing
Manufacturing, durables
Manufacturing, nondurables
Transportation, communication, and other public utilities (transportation and public utilities)
Trade (wholesale and retail)
Finance and service industries (meaning sv ind, 406, and tp26 ind codes 707-709)
Service industries (includes forestry and fisheries and private households)
Finance, insurance, and real estate
Service industries except private households (miscellaneous services on tabs, E.g., 55)
All nonagricultural industries except private households (this code has also been used in conjunction with 5:1 and 19:2 to identify
other nonagriculture private wage and salary than private households, e.g., tabs 15)
Also, long list of 3-digit SIC level industries.
Constraints or Interactions. To get choices of detailed SICs, status must equal civilian labor
force, other variables must equal null.
Notes.

2. Nonfarm Payroll Statistics from the Current Employment Statistics (National)
Industry-related variables and values.

industry About 19 screens of codes, division level, and 2-4-digit SIC. Additional combination values include

total non-farm, total private, goods-producing, wholesale and retail trade, private nonagricultural, manufacturing, major manufacturing, durable goods, nondurable goods, service producing industry, private service-producing industry.

Constraints or Interactions. Actual selection of industry values available depends on choice made for data type.

Notes.

3. Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)
Industry-related variables and values.

detail Total non-agricultural, 1-digit industry (Industry division), 2-digit industry
3-digit industry, A combination of industry divisions, A combination of industries, A residual cell

industry. All divisions, 2- and 3-digit SIC codes, plus

goods producing
service producing
construction and mining
services and mining (has asterisk)
Durable goods
Lumber & wood products including furniture
Prim metal inds & fabr. Metal prods
33-36
mach & electronic & electric equip.
nondurable goods
textile mill products & apparel
broadwoven and narrow fabric mills
men's & boys' suits, coats, & furngs.
Paper & printing & publishing
Pulp, paper and paperboard mills
261, 262, 263, 266
chemicals & petroleum & coal products
communications and public utilities
wholesale trade
retail trade
*R
Other durable goods
Other nondurable goods
Other federal government
Other state and local government

Constraints or Interactions. Value chosen for detail controls values displayed for industry. If seasonally adjusted, fewer values available.

Notes. Need explanation of detail value "a residual cell", and the industry value "*R".

4. Covered Employment and Wages
Industry-related variables and values.

size All

ownership Total covered, Federal government, State government, Local government, International government, Private

industry Division, 2-digit SIC, 3-digit SIC, 4-digit SIC, plus
Nonclassifiable establishments, All industries

Constraints or Interactions.

Notes. Explain current state of single value for size, expectations for additional values in future.

5. Mass Layoff Statistics
Industry-related variables and values.

IRC Total all industries, Total all industry, private, Agriculture, Nonagriculture, Manufacturing, Durable goods

(also list of specific durable goods industries, including "miscellaneous manufacturing industries"),

Nondurable goods,

(also list of specific nondurable goods industries),

Non manufacturing,

(also list of divisions),

Not identified, Government, Federal, State, Local,

Total, all reasons, Automation, bankruptcy, business ownership change, contract cancellation, contract completion, domestic relocation, energy-related, environment-related, financial difficulty, import competition, labor dispute, material shortage, model changeover, natural disaster, overseas relocation, plant or machine repair, product line discontinued, reorganization within company, seasonal work, slack work, vacation period, weather-related, other reasons, reason not reported

Characteristics = black, Hispanic, women

Constraints or Interactions.

Notes. IRC contains many kinds of information, including ownership as well as industry type. Also includes non-industry-related choices.

6. Producer Price Index Revision – Current Series
Industry-related variables and values.
industry 2-digit SIC Codes, 3-digit SIC Codes, 4-digit SIC Codes,

Crude processors
Crude processors
Crude processors
Finished processors
Primary processors
Primary processors
Semifinished processors
Non-residential buildings
Inputs to construction industries
Final demand
Final demand
Finished processors
Capital investment
Other heavy construction
Highway and street construction
Non-residential
Maintenance and repair construction
Residential
New construction
Consumers
Consumers
Primary processors
Single-unit residential
Multi-unit residential
Semifinished processors
Water and sewer supply construction
Total manufacturing industries
Total mining industries

product Long list of product codes, organized by industry code (including non-SIC) categories. Also includes combinations such as durables, nondurables, goods less food and energy, etc.

Constraints or Interactions. Choice of industry value constrains list of product values.

Notes.

7. Producer Price Index Revision – Discontinued Series
Note: same as current series above.
Industry-related variables and values.
Constraints or Interactions.
Notes.

8. Collective Bargaining Statistics – Historical (Private Sector)
Industry-related variables and values.
industry All industries, Manufacturing, Nonmanufacturing, Construction, Goods-producing, Service-producing
Constraints or Interactions.
Notes.

9. Collective Bargaining Statistics – Historical (State and Local Government)
Industry-related variables and values.

<u>sector</u> State and local government. State government, Local government
Constraints or Interactions.
Notes.

10. Employee Benefits Survey
Industry-related variables and values.

<u>type</u> medium and large private sector, state and local governments, small private sector establishments
Constraints or Interactions.
Notes. Note combination of size and sector.

11. Employment Cost Trends (Employment Cost Index)
Industry-related variables and values.

<u>group</u> values including industry are shown below. Others include occupation, geographical region, union status.

Wholesale and retail trade; excluding sales occupations
Finance, insurance, and real estate; excluding sales occupations
Wholesale trade; excluding sales occupations
Goods-producing industries
Goods-producing industries; excluding sales occupations
Goods-producing industries; White-collar occupations
Goods-producing industries; White-collar occupations, excluding sales occupations
Goods-producing industries; Blue-collar occupations
Goods-producing industries; Service occupations
Service-producing industries
Service-producing industries; excluding sales occupations
Service-producing industries; White-collar occupations
Service-producing industries; White-collar occupations, excluding sales occupations
Service-producing industries; Blue-collar occupations
Service-producing industries; Service occupations
Non-manufacturing industries
Non-manufacturing industries; White-collar occupations
Non-manufacturing industries; White-collar occupations, excluding sales occupations
Non-manufacturing industries; Blue-collar occupations
Non-manufacturing industries; Service occupations
Construction
Manufacturing
Manufacturing - durable goods
Manufacturing - non-durable goods
Manufacturing; White-collar occupations
Manufacturing; White-collar occupations, excluding sales occupations
Manufacturing; Blue-collar occupations
Manufacturing; Service occupations
Transportation and Public Utilities
Transportation
Public utilities
Communications
Electric, gas, and sanitary services
Wholesale and retail trade
Wholesale trade
Retail trade
General merchandise stores
Food stores
Finance, insurance, and real estate
Banking, savings and loan, and other credit agencies
Insurance carriers, agents, brokers, and service
Insurance, excluding sales occupations
Services industries
Schools
Elementary and secondary schools
Services industries, excluding schools
Health services

Appendix A   11

Hospitals
Nursing and personal care facilities
Business services
Educational services
Public administration
Colleges and universities
Union; Manufacturing; Blue-collar occupations
Union; Manufacturing
Union; Non-manufacturing industries
Union; Goods-producing industries
Union; Service-producing industries
Non-union; Manufacturing; Blue-collar occupations
Non-union; Manufacturing
Non-union; Non-manufacturing
Non-union; Goods-producing industries
Non-union; Service-producing industries
Aircraft and parts manufacturing (SIC 372)
Aircraft and parts manufacturing (SIC 372); White-collar occupations
Aircraft and parts manufacturing (SIC 372);Professional, specialty, and technical occupations
Aircraft and parts manufacturing (SIC 372); Executive, administrative, and managerial occupations
Aircraft and parts manufacturing (SIC 372);Administrative support, including clerical, occupations
Aircraft and parts manufacturing (SIC 372); Blue-collar occupations
Aircraft and parts manufacturing (SIC 372); Precision, production, craft, and repair occupations
Aircraft and parts manufacturing (SIC 372); Machine operators, assemblers, and inspectors occupations
Aircraft manufacturing (SIC 3721)
Aircraft manufacturing (SIC 3721); White-collar occupations
Aircraft manufacturing (SIC 3721); Blue-collar occupations
Aircraft engines and engine parts (SIC 3724)
Aircraft engines and engine parts (SIC 3724); White-collar occupations
Aircraft engines and engine parts (SIC 3724); Blue-collar occupations
Aircraft parts and equipment, NEC (SIC 3728)
Aircraft parts and equipment, NEC (SIC 3728); White-collar occupations
Aircraft parts and equipment, NEC (SIC 3728); Blue-collar occupations
Guided missiles and space vehicles manufacturing (SIC 3761)
Guided missiles and space vehicles manufacturing (SIC 3761); White-collar occupations
Guided missiles and space vehicles manufacturing (SIC 3761); Blue-collar occupations

ownership Civilian, Private industry, State and local government

Constraints or Interactions. Values displayed for group depend on choices for seasonal adjustment, compensation, and ownership.

Notes.

12. Nonfatal Cases Involving Days Away From Work: Selected Characteristics
Industry-related variables and values.
case type
Industry division or selected characteristic by detailed nature of condition
Industry division or selected characteristic by detailed part of body affected
Industry division or selected characteristic by detailed source of injury/illness
Industry division or selected characteristic by detailed event or exposure
Industry division or selected characteristic by detailed occupation
Selected injury/illness characteristic by detailed industry
Industry division or length of absence by age group
Industry division or length of absence by gender
Industry division or length of absence by race
Industry division or length of absence by length of service
category
Total private industry
Industry division – agriculture
Industry division – mining
Industry division – construction
Industry division – manufacturing
Industry division – transportation and public utilities
Industry division – wholesale trade
Industry division -- retail trade
Industry division – finance, insurance, and real estate
Industry division – services

Also includes values for age groups, days away from work, event or exposure, gender, length of service, nature of injury, occupation, part of body affected, race, source of injury.

industry Agriculture, forestry, and fishing, Mining, Construction, Manufacturing, Transportation, communications, electric, gas, and sanitary services, Wholesale trade, Retail trade, Finance, insurance, and real estate, Services,

Plus 2- and 3-digit SIC codes.

Constraints or Interactions. Detailed industry values available only if case type = detailed industry.

Notes. The order in which occupations are listed seems to be based on codes, which results in some very odd sequences. Can this be displayed as a drill-down list?

13. Census of Fatal Occupational Injuries
Industry-related variables and values.

case type Fatalities by detailed industry (all sectors), Fatalities by detailed private industry, Fatalities by detailed government industry

Also has event, occupation, source of injury values.

category
All industries
All private industry
Private industry division – agriculture
Private industry division – mining
Private industry division – construction
Private industry division – manufacturing
Private industry division – transportation and public utilities
Private industry division – wholesale trade
Private industry division – retail trade
Private industry division – finance, insurance, and real estate
Private industry division – services
All government
Federal government
State government
Local government

Also includes values for age, gender, occupation, worker location, worker activity, primary and secondary source of injury.

industry All workers
Industry divisions
2- and 3-digit industries

Constraints or Interactions. If case type is a value that doesn't contain "detailed industry", the industry division choices are presented in category. Otherwise, all values for industry are available.

Notes. The order in which occupations are listed seems to be based on codes, which results in some very odd sequences. Can this be displayed as a drill-down list?

14. Occupational Injuries and Illnesses: Industry Data (pre 1989)
Industry-related variables and values.

division
Private industry
Agriculture, forestry, and fishing
Mining
Construction
Manufacturing
Transportation and public utilities
Wholesale and retail trade
Wholesale trade
Retail trade
Finance, insurance, and real estate

Services

industry 2-,3-,4(?) digit SIC codes within divisions

Constraints or Interactions. Choice of division controls choices of industry; only those within chosen division are shown.

Notes.

15. Occupational Injuries and Illnesses: Industry Data (1989-current)

Industry-related variables and values.

<u>division</u>

Private industry
Agriculture, forestry, and fishing
Mining
Construction
Manufacturing
Transportation and public utilities
Wholesale and retail trade
Wholesale trade
Retail trade
Finance, insurance, and real estate
Services

industry 2-,3-,4(?) digit SIC codes within divisions

Constraints or Interactions. Choice of division controls choices of industry; only those within chosen division are shown.

Notes.

16. Major Sector Productivity and Costs Index

Industry-related variables and values.

<u>sector</u>

Manufacturing
Manufacturing, Durable Goods
Manufacturing, Nondurable Goods
Business
Nonfarm Business
Farm Business
Nonfinancial Corporations

Constraints or Interactions.

Notes. Notice that sector here means industry type, not ownership.

17. Major Sector Multifactor Productivity Index

Industry-related variables and values.

<u>sector</u>

Manufacturing
Nondurable goods (SIC 20-23, 26-31)
Food & kindred prod. (SIC 20)
Textile mills prod (SIC 22)
Apparel & related prod. (SIC 23)
Paper & allied prod. (SIC 26)
Printing & publishing (SIC 27)
Chem. & allied prod (SIC 2)
Petroleum refining (SIC 29)
Rubber & plastic prod. (SIC 30)
Durable good (SIC 24-25, 32-39)
Lumber & wood prod (SIC 24)
Furniture & fixtures (SIC 25)
Stone, clay & glass (SIC 32)
Primary metal ind. (SIC 33)
Fabricated metal prod. (SIC 34)
Ind. Machinery, comp. eq. (SIC 35)
Electric & electr. eq. (SIC 36)
Transportation equip. (SIC 37)
Instruments (SIC 38)
Misc. manufacturing (SIC 39)

Elect & gas utilities (SIC 49)
Private business
Private nonfarm business

Constraints or Interactions.

Notes. Notice that sector here means industry type, not ownership.

18. International Labor Statistics

Industry-related variables and values.

group

MANUFACTURING OUTPUT INDEX
MFG AGGREGATE HOURS INDEX
MFG AGGREGATE COMPENSATION INDEX
MFG OUTPUT PER HOUR INDEX
MFG HR COMP INDEX, NATL CURRENCY BASIS
MFG HR COMP INDEX, US DOLLAR BASIS
MFG UNIT LABOR COST INDEX, NATL CURRENCY BASIS
MFG UNIT LABOR COST INDEX, US DOLLAR BASIS
MANUFACTURING EMPLOYMENT INDEX
MFG AVERAGE HOURS INDEX

Plus other indexes related to CPI, employment, etc.

Constraints or Interactions. Choice of country limits indexes available.

Notes.

Extended Example II: The Establishment Size Concept

The concept of *establishment size* as used in the BLS is based on the establishment's number of employees (per EBS). However, users may have different measures of size in mind, e.g., number of employees, annual profits, number of products, etc. Also note the possible confusion between industry size and establishment size. This concept was included first, because of its relationship to the major concept *industry*, and second, because its associated variables and values demonstrate some of the problems in "nailing down" the definition for a somewhat vague idea.

A. User Vocabulary Examples.
Many of the nouns in the vocabulary list contain synonyms for establishment.

Company size, corporation size, employer size, firm size, retail giant, small business, small-to-midsize companies, microenterprise, fast-growing business, major employer.

B. BLS Concept
Establishment size
See also the concept of *industry*. Consider also pointing to resources that are based on different measures of size.

C. Surveys/Series
1. Covered Employment and Wages
2. Employee Benefits Survey

D. Word-Concept Aids
1. General definition of *establishment size* as used in the BLS, i.e., that is based on number of employees. Employee Benefits Survey defines "small" as < 100 employees, and medium and large as >+ 100 employees. Is this definition used throughout BLS?

2. Distinction between establishment size and industry size. (See introduction to this example.)

3. Synonyms and related terms within the BLS, if any, such as *number of workers*, or *number of employees*.

4. Parsing possibilities.
A noun phrase template along with a lexicon of words for each "part of speech" (i.e., SIZE, ESTABLISHMENT) may identify a good portion of queries containing a specification of establishment size, but this would need to be tested to be sure there wouldn't be too many false drops.
> Template: SIZE + ESTABLISHMENT
> Lexicon:
>> SIZE ->size, small, midsize, medium, large, giant.
>> Note: growing and shrinking also refer to size, but imply comparison of an establishment over time.
>> ESTABLISHMENT -> establishment, company, firm, business, organization, employer.
>> See also the Establishment concept in Appendix C.

5. Example questions, either illustrating frequently asked questions (FAQs), or to demonstrate how variables/values can be combined.

How many small businesses provide child care?
How much vacation do government employees receive after ten years' service?

E. Concept-Survey/Series Aids
For each relevant survey/series (see C):
- A 1-2 sentence summary of the information available in the survey/series.
- Definition and/or scope of industry concept as used in the survey/series, especially noting any restrictions in coverage.
- List of all variables available in the survey/series, with definition where necessary. (E.g., from Series ID Formats, http://stats.bls.gov/hlpforma.htm.)
- Highlight those variables related to industry.

Also list any possible sources of related data, in case these resources can't satisfy the user's information need.

1. Covered Employment and Wages
Summary. Includes data on the number of establishments, monthly employment, and quarterly wages, by industry, at the 4-digit Standard Industrial Classification (SIC) level, by county, by ownership sector, for the entire United States. These data are aggregated to annual levels, to higher industry levels (3-digit, 2-digit, industry division and totals), and to higher geographic levels (national, State, and Metropolitan Statistical Area (MSA))
Definition and Scope. At the moment, there is only one value for size, but this is in the works.
Variables. Seasonal adjustment, area, datatype, ownership, industry,
     size - (of establishment).
Other Sources.

2. Employee Benefits Survey
Summary. The Employee Benefits Survey (EBS) of the Bureau of Labor Statistics (BLS) covers the incidence and characteristics of employee benefit plans.
Definition and Scope. Size values only apply to private sector establishment.
Variables. Seasonal adjustment, title (category of benefit provision),
    type - (size and type of establishment).
Other Sources.

F. Survey/Series – Query Aids
- For each relevant survey/series (see C):
- Lists of values for establishment size-related variables. Note: literal values are in small font, descriptions of values are in normal size.
- Description of constraints or interactions between establishment size-related variable-value choices.
- Notes on possible issues with results, e.g., breaks in collection, preliminary data, changes in definition.

1. Covered Employment and Wages
Establishment size-related variables and values.
size all
Constraints or Interactions.
Notes. Need to explain schedule for providing information broken out by size.

2. Employee Benefits Survey
Establishment size-related variables and values.

<u>type</u> Medium and large private sector, State and local governments, Small private sector establishments

Constraints or Interactions.

Notes. Documentation for the survey defines small as < 100 employees, and medium/large as >= 100 employees.

Extended Example III: The Occupation Concept

This concept describes the type of job a person has, or work in which he/she engages as a whole, (e.g., "give me unemployment figures for January 2001 by occupation") and specific occupation types, (e.g., "How many people are employed as realtors"). This topic was chosen in part because of the frequency with which it appears in user questions. Related concepts are given in section B.

A. User Vocabulary Examples
job, career, occupation, work, business (ambiguous with a specific type of occupation), profession, trade, job title, job description, duties
Specific occupation names, such as, accountant, bus-driver, farmer, housekeeper, nurse, plumber, teacher. Note the varying levels of specificity, e.g., types of nurses.

B. BLS Concept
Occupation
See also the concepts of white collar/blue collar, self-employed, work at home.

C. Surveys/Series
1. Labor Force Statistics from the Current Population Survey
2. Geographic Profile
3. Consumer Expenditure Survey
4. National Compensation Survey
5. Employment Cost Index – Seasonal (Employment Cost Trends)
6. Nonfatal Cases Involving Days Away From Work: Selected Characteristics
7. Census of Fatal Occupational Injuries

D. Word-Concept Aids
1. General definition of *occupation* as used in the BLS. (Drawn from BLS document.)

2. Distinction between industry and occupation. An industry is a public or private organization or institution that … An occupation is the job or career that an individual follows, usually in return for compensation of some sort. In general language, the same word is often used for both, e.g., *farming*, or *plumbing*. In the BLS data, however, there is a distinction between the industry of *agriculture* and the occupation of *farming*, and a similar distinction between *plumbing* and *plumbers*. (Also use this clarification of the *industry* concept.)

3. Synonyms and related terms within the BLS. See, for example, the variable names used within the series, and those used for the see-also concepts listed in B.
class of worker

4. Link to Standard Occupational Classification, along with a brief description of its structure and use. http://www.bls.gov/soc/socguide.htm

5. Examples of SOC 1) major group; 2) minor group; 3) broad occupation; and 4) detailed occupation coding, to give users the idea of varying levels of specificity.
a. Major Group 15-0000 Computer and Mathematical Occupations
   Minor Group 15-1000 Computer Specialist
      Broad Occupation 15-1030 Computer Software Engineers

Detailed Occupation 15-1031 Computer Software Engineers, Applications
b. Major Group 31-0000 Healthcare Support Occupations
    Minor Group 31-9000 Other Healthcare Support Occupations
        Broad Occupation 31-9090 Miscellaneous Healthcare Support Occupations
        Detailed Occupation 31-9091 Dental Assistants
c. Major Group 41-0000 Sales and Related Occupations
    Minor Group 41-9000 Other Sales and Related Workers
        Broad Occupation 41-9020 Real Estate Brokers and Sales Agents
        Detailed Occupation 41-9022 Real Estate Sales Agents

6. Link to Occupational Outlook Handbook, along with a brief description of its structure and use.
http://www.bls.gov/ocohome.htm.

7. Link to Career Guide to Industry, along with a brief description of its structure and use.
http://www.bls.gov/cghome.htm.

8. Link to O*Net Online, which is a source of job titles and descriptions.
http://online.onetcenter.org

9. Parsing possibilities.
- For general occupation words listed in A, show user descriptions of surveys/series listed in E2. Or add an intermediate step, first allowing user to choose occupation concept or one of the see-also concepts listed in B. (Use similar treatment for *industry*.)
- Always have options of browsing SOC available to user. It is not reasonable to maintain a lexicon of all possible words naming specific industries or occupations.

10. Example questions, either illustrating frequently asked questions (FAQs), or to demonstrate how variables/values can be combined.
    How much do high school teachers in North Carolina earn?

E. Concept-Survey/Series Aids
For each relevant survey/series (see C):
- A 1-2 sentence summary of the information available in the survey/series.
- Definition and/or scope of occupation concept as used in the survey/series, especially noting any restrictions in coverage.
- List of all variables available in the survey/series, with definition where necessary. (E.g., from Series ID Formats, http://stats.bls.gov/hlpforma.htm.)
- Highlight those variables related to occupation.

Also list any possible sources of related data, in case these resources can't satisfy the user's information need.

1. Labor Force Statistics from the Current Population Survey
Summary. This series presents information about the size and makeup of the U.S. labor force, which can be broken out by several demographic variables.
Definition and Scope. Occupation covers all occupations at various levels of specificity. Can also get information on unpaid family workers and self-employed through the class of worker variable.
Variables. Seasonal adjustment, age, sex, race, ethnicity, status (portion of labor force included), industry,

       class of worker - (sector in which individual works)

       occupation - major occupational groups and subgroups.

Other Sources.

2. Geographic Profile

Summary. Based on Local Area Unemployment Statistics, this presents occupation, demographic, and employment data for a variety of geographic areas.

Definition and Scope. Occupation is defined at the major group level.

Variables. Seasonal adjustment, state, area, labor force (includes employed/unemployed),

       character code - (combines occupation values with age, ethnicity, race, and sex).

Other Sources.

3. Consumer Expenditure Survey

Summary. This survey provides information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics.

Definition and Scope. Occupation is that of the reference person of survey, broken out by very broad categories, including retired and not reporting.

Variables. Item (demographics of reference person or household, and consumer items),

       table - ("standard characteristics" of how data is broken out, occupation is one of these),

       column - (subsets of major characteristics – lists occupation values if table = occupation).

Other Sources.

4. National Compensation Survey

Summary. The NCS covers civilian workers in private industry establishments and State and local governments. It excludes private households, the Federal government, and agriculture. It provides wage and benefit data for localities, geographic regions, and the U.S. as a whole.

Definition and Scope. Occupation excludes private households and agriculture.

Variables. State (including all U.S.), area (within a single state),

       occupation - Many choices of general and specific occupation groups.

       level - Based on duties performed in the job, allows comparison across occupations.

Other Sources.

5. Employment Cost Index – Seasonal (Employment Cost Trends)

Summary. The ECI measures the change over time in the cost of labor. The cost of labor includes wages and salaries and employer costs for employee benefits.

Definition and Scope. Occupation variables are broad, often at the blue/white collar level.

Variables. Seasonal adjustment, compensation (wage, benefits, or both), periodicity (quarterly or annual average), ownership (sector of industry owner),

       group - (industry, occupation, region, union status of compensation – some in

       combination).

Other Sources.

6. Nonfatal Cases Involving Days Away From Work: Selected Characteristics

Summary. Since 1972, the survey has reported annually on the number of workplace injuries and illnesses in private industry and the frequency of those incidents.

Definition and Scope.

Variables. Seasonal adjustment, data type (rate of injury/illness, number of cases, or median workdays lost),

       case type - (how data is broken out, including by detailed occupation),

category - (identifies the major occupation group, injury and illness classification, industry, or demographic characteristic. A complex combination of values), occupation - (detailed list of occupations).
Other Sources.

7. Census of Fatal Occupational Injuries
Summary. Provides information on 28 separate data elements including information on the worker, the fatal incident, and the machinery or equipment involved.
Definition and Scope.
Variables. Data type (number of fatal injuries), case type - (how data is broken out, including occupation, industry, event, source of injury), category - (identifies the broad occupation group, industry, injury and illness classification, or demographic characteristic. A complex combination of values), occupation - (detailed list of occupations).
Other Sources.

F. Survey/Series – Query Aids
- For each relevant survey/series (see C):
- Lists of values for occupation-related variables. Note: literal values are in small font, descriptions of values are in normal size.
- Description of constraints or interactions between occupation-related variable-value choices.
- Notes on possible issues with results, e.g., breaks in collection, preliminary data, changes in definition.

1. Labor Force Statistics from the Current Population Survey
Occupation-related variables and values.
class of worker N/A, Wage and Salary Workers, Private Wage and Salary Workers, Government Wage and Salary Workers, Self-employed Workers, Unpaid Family Workers, Self-employed And Unpaid Family Workers (8 and 9)

occupation n/a, managerial & prof. Specialty; executive, admin, & managerial; officials & admin pub, admin; other exec, admin, & managerial; management related occupations; professional specialty occupations; engineers; math & computer scientists; natural scientist; health diagnosing occupations; health assessment & treating; teachers, college & univ; teachers, exc college & univ; lawyers & judges; other prof, specialty occ; technical, sales, admin support; technicians & related support occ; health technologists & science techs; engineering, & science tech; techs, except health, eng., sci; sales occupations, supervisors & proprietors; sales reps, finance & bus serv; sales reps, commodities ex retail; sales, retail and personal; sales related occupations; admin support, inc clerical; supervisors; computer equipment operators; secs, stenos, & typists; financial records processing; mail & message distributing; oth. Admin support, inc clerical; service occupation; private household; protective services; service, exc pvt hhld & protective; food service; health service; cleaning & building service; persona service; precision production, craft, repair; mechanics & repairers; construction trades; other production, craft, repair; operators, fabricators, laborers; mach opts, asmblrs, inspectors occupations; operators & tenders, exc prec; fbrctrs, asmblrs, inspct, smplrs industries; transportation, material moving; motor vehicle operators; other trans & material occ; hndlrs, equip clnrs, hlpers, laborers; construction laborers; other hndlrs, equip clners, hlpers, laborers; freight, stock mat handlers, other handlers exc freight, stock, & mat handlers; farming forestry, fishing, farm operators & managers; farm workers & related occ; forestry and fishing occupations; other farming, forestry and fishing occ; aggregated total; all nonfarm occupations

Constraints or Interactions. A smaller group of high-level occupations are presented with seasonal adjustment; detailed list available without seasonal adjustment. Choices for demographic variables also limit list of occupations.
Notes.

2. Geographic Profile
Occupation-related variables and values.
character code Occupation values are combined with age, race, ethnicity, and sex values.
executive, administrative, and managerial;
 professional specialty;
technicians and related support;
sales;

administrative support, including clerical;
service occupations;
precision production, craft, and repair;
machine operators, assemblers, and inspectors;
transportation and material moving;
farming, forestry, and fishing
Constraints or Interactions.
Notes.

3. Consumer Expenditure Survey
Occupation-related variables and values.

<u>table</u> quintiles of income, income before taxes, age of reference person, size of consumer unit, composition of consumer unit, number of earners in consumer unit,
housing tenure, race of reference person, type of area, region of residence, occupation of reference person, and origin of reference person.

<u>column</u>
Self-employed workers
Total wage and salary earners
Wage and salary earners: managers and professionals
Wage & salary: tech., sales, and clerical workers
Wage & salary earners; service workers
Wage & salary: construction workers & mechanics
Wage & salary: operators, fabricators & laborers
Retired
Occupation: all other, including not reporting
Constraints or Interactions. Table value must = occupation of reference person.
Notes.

4. National Compensation Survey
Occupation-related variables and values.
<u>occupation</u> Broad and specific occupations are also listed within the major and minor groups included here.
All occupations
All occupations, excluding sales
White collar occupations
    White collar occupations, excluding sales
        Professional specialty and technical occupations
        Engineers, architects, and surveyors
        Mathematical and computer scientists
        Natural scientists
        Health related occupations
        Teachers, college and university
        Teachers, except college and universities
        Librarians, archivists and curators
        Social scientists and urban planners
        Social, recreation, and religious workers
        Lawyers and judges
        Writers, authors, entertainers, athletes and professionals
        Technical occupations
        Executive, administrative, and managerial occupations
        Executives, administrators and mangers
        Management related occupations
        Sales occupations
        Administrative support occupations, including clerical
Blue collar occupations
        Precisions production, craft, and repair occupations
        Machine operators, assemblers, and inspectors
        Transportation and material moving occupations
        Handlers, equipment cleaners, helpers, and laborers
        Service occupations, except private household
        Protective service occupations
        Food service occupations
        Health service occupations
        Cleaning and building service occupations
Personal service occupations

<u>level</u> Levels 01 - 15
Overall occupation average (no work level)
Not able to be leveled

Constraints or Interactions. Levels are only applicable on the broad occupation level.

Notes. Note the use of "blue collar" and "white collar". Can these listings be considered official definitions?

5. Employment Cost Index – Seasonal (Employment Cost Trends)

Occupation-related variables and values.

<u>group</u> group values including occupation are shown below. Others include industry, geographical region, union status.

All workers
Production and non-supervisory occupations
All workers, excluding sales occupations
White-collar occupations, excluding sales occupations
White-collar occupations
Executive, administrative, and managerial occupations
Professional, specialty, and technical occupations
Sales occupations
Administrative support, including clerical, occupations
Blue-collar occupations
Precision, production, craft, and repair occupations
Machine operators, assemblers, and inspectors occupations
Transportation and material moving occupations
Handlers, equipment cleaners, helpers, and laborers occupations
Service occupations
All workers, excluding sales occupations
White-collar occupations, excluding sales occupations
Wholesale and retail trade; excluding sales occupations
Finance, insurance, and real estate; excluding sales occupations
Wholesale trade; excluding sales occupations
Goods-producing industries; excluding sales occupations
Goods-producing industries; White-collar occupations
Goods-producing industries; White-collar occupations, excluding sales occupations
Goods-producing industries; Blue-collar occupations
Goods-producing industries; Service occupations
Service-producing industries; excluding sales occupations
Service-producing industries; White-collar occupations
Service-producing industries; White-collar occupations, excluding sales occupations
Service-producing industries; Blue-collar occupations
Service-producing industries; Service occupations
Non-manufacturing industries; White-collar occupations
Non-manufacturing industries; White-collar occupations, excluding sales occupations
Non-manufacturing industries; Blue-collar occupations
Non-manufacturing industries; Service occupations
Manufacturing; White-collar occupations
Manufacturing; White-collar occupations, excluding sales occupations
Manufacturing; Blue-collar occupations
Manufacturing; Service occupations
Insurance, excluding sales occupations
Union; Blue-collar occupations
Union; Manufacturing; Blue-collar occupations
Non-union; Blue-collar occupations
Non-union; Manufacturing; Blue-collar occupations
Aircraft and parts manufacturing (SIC 372); White-collar occupations
Aircraft and parts manufacturing (SIC 372);Professional, specialty, and technical occupations
Aircraft and parts manufacturing (SIC 372); Executive, administrative, and managerial occupations
Aircraft and parts manufacturing (SIC 372);Administrative support, including clerical, occupations
Aircraft and parts manufacturing (SIC 372); Blue-collar occupations
Aircraft and parts manufacturing (SIC 372); Precision, production, craft, and repair occupations
Aircraft and parts manufacturing (SIC 372); Machine operators, assemblers, and inspectors occupations
Aircraft manufacturing (SIC 3721); White-collar occupations
Aircraft manufacturing (SIC 3721); Blue-collar occupations
Aircraft engines and engine parts (SIC 3724); White-collar occupations
Aircraft engines and engine parts (SIC 3724); Blue-collar occupations
Aircraft parts and equipment, NEC (SIC 3728); White-collar occupations
Aircraft parts and equipment, NEC (SIC 3728); Blue-collar occupations

Appendix A   24

Guided missiles and space vehicles manufacturing (SIC 3761); White-collar occupations
Guided missiles and space vehicles manufacturing (SIC 3761); Blue-collar occupations
Constraints or Interactions.
Notes. Note the frequent use of blue/white collar. Clarify that these follow NCS definitions?

6. Nonfatal Cases Involving Days Away From Work: Selected Characteristics
Occupation-related variables and values.
case type
Industry division or selected characteristic by detailed nature of condition
Industry division or selected characteristic by detailed part of body affected
Industry division or selected characteristic by detailed source of injury/illness
Industry division or selected characteristic by detailed event or exposure
Industry division or selected characteristic by detailed occupation
Selected injury/illness characteristic by detailed industry
Industry division or length of absence by age group
Industry division or length of absence by gender
Industry division or length of absence by race
Industry division or length of absence by length of service
category
Executive and professional specialty
Technical, sales and administrative support
Service
Farming
Precision production
Operators, fabricators
Also includes values for age groups, days away from work, event or exposure, gender, length of service, nature of injury, industry division, part of body affected, race, source of injury.
occupation long list of detailed occupations.
Constraints or Interactions. Occupation choices are only available if datatype is # cases or median days lost, and case type is by detailed occupation.
Notes. The order in which occupations are listed seems to be based on codes, which results in some very odd sequences. Can this be displayed as a drill-down list?

7. Census of Fatal Occupational Injuries
Occupation-related variables and values.
case type Fatalities by detailed occupation.
Also has industry, event, source of injury values.
category
Executive and professional specialty
Technical, sales and administrative support
Sales
Service Protective service
Farming
Precision production
Construction
Operators, fabricators
Material movement
Handlers and laborers
Military
Also includes values for age, gender, industry, worker location, worker activity, primary and secondary source of injury.
occupation Long list of detailed occupations, as in #6 above.
Constraints or Interactions. Occupation choices are only available if case type is by detailed occupation.
Notes. The order in which occupations are listed seems to be based on codes, which results in some very odd sequences. Can this be displayed as a drill-down list?

Extended Example IV: White Collar/Blue Collar

This is a somewhat old-fashioned-sounding pair of terms that classifies occupations roughly into production/manual labor, and professional/desk jobs.

- Liddy & Liddy (2001) classified occupations into *white collar*, *blue collar*, and *pink collar*. Standard dictionaries generally define *white collar* as professional and/or desk jobs, and *blue collar* as including manual labor of some kind. The origin of the term lies in the idea that since this kind of work tends to be messy, a colored shirt would show the dirt less than a white one. *Pink collar* refers to low-paying jobs traditionally held by women, and is therefore generally considered nonpreferred English.
- In an entirely unscientific sample of a handful of graduate students, they all said that had used *white collar*. They had heard of *blue collar*, but said they really wouldn't feel comfortable using it to describe anyone's job currently, although it would be ok in a historical context. Most of them had not heard the term *pink collar*, and when it was explained to them, said that they would use it only in a historical context, if at all.
- The Occupational Outlook Handbook (http://www.bls.gov/ocohome.htm) uses *blue collar* only in the job title "blue collar worker supervisor". It uses *white collar* only in the description of typical duties for police and detectives – dealing with white collar crime.
- The NCS seems to classify occupations, thus providing an "official" BLS definition.

It should be noted that these are fuzzier concepts in general language. This is a good example of cases where the user may not have a crisp, iron-clad definition in terms of included/excluded occupations, whereas the BLS does.

A. User Vocabulary Examples.
White collar, white collar worker, desk worker, professional, non-manual worker,
Blue collar, blue collar worker, production worker, unskilled worker, laborer

B. BLS Concept
White collar occupations, blue collar occupations
See also the concept of *occupation*.

C. Surveys/Series
Note: The ones listed here are those that explicitly use these terms as values.
1. National Compensation Survey
2. Employment Cost Index – Seasonal (Employment Cost Trends)

D. Word-Concept Aids
1. Definitions via reference to SOC categories for each term.

2. The term *pink collar* has also been used for traditional female jobs such as secretary or clerical work. To my knowledge, this term is not used by the BLS, but an explicit disclaimer to this effect may be useful. In searching the OOH, I found 1 job title that used *blue collar* (Blue collar worker supervisors). The only use of *white collar* I found referred to "white collar crime" in describing the duties of Police and Detectives.

3. *Laborer* and *production worker* are probably more specific terms than blue collar, but users may consider them to synonyms. Ditto *professional* and *white collar*.

4. Parsing possibilities.
- Use of these terms indicates a group of occupations, so the pertinent surveys/series could be displayed immediately.
- It is not necessarily the case that the user's definition of them matches how they are used in the series listed in C. These occupations may be highlighted, but shouldn't necessarily be assumed to match those of interest.

5. Example questions, either illustrating frequently asked questions (FAQs), or to demonstrate how variables/values can be combined.

> What is the average pay of a blue-collar worker in Alabama?

E. Concept-Survey/Series Aids
For each relevant survey/series (see C):
- A 1-2 sentence summary of the information available in the survey/series.
- Definition and/or scope of *white collar/blue collar* concept as used in the survey/series, especially noting any restrictions in coverage.
- List of all variables available in the survey/series, with definition where necessary. (E.g., from Series ID Formats, http://stats.bls.gov/hlpforma.htm.)
- Highlight those variables related to industry.

Also list any possible sources of related data, in case these resources can't satisfy the user's information need.

1. National Compensation Survey
Summary. The NCS covers civilian workers in private industry establishments and State and local governments. It excludes private households, the Federal government, and agriculture. It provides wage and benefit data for localities, geographic regions, and the U.S. as a whole.
Definition and Scope. Occupation excludes private households and agriculture. Detailed occupations are implicitly classified as BC/WC by placement in list. The terms themselves are also used as values.
Variables. State (including all U.S.), area (within a single state), level (based on duties performed in the job, allows comparison across occupations),
> occupation - Many choices of general and specific occupation groups.

Other Sources.

2. Employment Cost Index – Seasonal (Employment Cost Trends)
Summary. The ECI measures the change over time in the cost of labor. The cost of labor includes wages and salaries and employer costs for employee benefits.
Definition and Scope. Occupation variables are broad, often at the blue/white collar level
Variables. Seasonal adjustment, compensation (wage, benefits, or both), periodicity (quarterly or annual average), ownership (sector of industry owner),
> group - (industry, occupation, region, union status of compensation – some in combination).

Other Sources.

F. Survey/Series- Query Aids
For each relevant survey/series (see C):
- Lists of values for industry-related variables. Note; literal values are in small font, descriptions of values are in normal size.
- Description of constraints or interactions between industry-related variable-value choices.

- Notes on possible issues with results, e.g., breaks in collection, preliminary data, changes in definition.

1. National Compensation Survey
BC/WC-related variables and values (i.e., those that explicitly use the term).
occupation Broad and specific occupations are also listed within the major and minor groups included here. See NCS #4 in Extended Example III, Occupation.
White collar occupations
White collar occupations, excluding sales
Blue collar occupations
Constraints or Interactions.
Notes.

2. Employment Cost Index – Seasonal (Employment Cost Trends)
BC/WC-related variables and values (i.e., those that explicitly use the term).
group group values including bc/wc are shown below. Others include additional occupations, industry, geographical region, union status.
White-collar occupations, excluding sales occupations
White-collar occupations
Blue-collar occupations
White-collar occupations, excluding sales occupations
Goods-producing industries; White-collar occupations
Goods-producing industries; White-collar occupations, excluding sales occupations
Goods-producing industries; Blue-collar occupations
Service-producing industries; White-collar occupations
Service-producing industries; White-collar occupations, excluding sales occupations
Service-producing industries; Blue-collar occupations
Non-manufacturing industries; White-collar occupations
Non-manufacturing industries; White-collar occupations, excluding sales occupations
Non-manufacturing industries; Blue-collar occupations
Manufacturing; White-collar occupations
Manufacturing; White-collar occupations, excluding sales occupations
Manufacturing; Blue-collar occupations
Union; Blue-collar occupations
Union; Manufacturing; Blue-collar occupations
Non-union; Blue-collar occupations
Non-union; Manufacturing; Blue-collar occupations
Aircraft and parts manufacturing (SIC 372); White-collar occupations
Aircraft and parts manufacturing (SIC 372); Blue-collar occupations
Aircraft manufacturing (SIC 3721); White-collar occupations
Aircraft manufacturing (SIC 3721); Blue-collar occupations
Aircraft engines and engine parts (SIC 3724); White-collar occupations
Aircraft engines and engine parts (SIC 3724); Blue-collar occupations
Aircraft parts and equipment, NEC (SIC 3728); White-collar occupations
Aircraft parts and equipment, NEC (SIC 3728); Blue-collar occupations
Guided missiles and space vehicles manufacturing (SIC 3761); White-collar occupations
Guided missiles and space vehicles manufacturing (SIC 3761); Blue-collar occupations
Constraints or Interactions.
Notes.

Appendix B

I. Industry and Related Variables
II. Occupation and Related Variables

Background information showing the LABSTAT surveys/series and variables in which the *industry* and *occupation* concepts occur.

I. Industry and Related Variables, Selective Access

Series	Variable Name	Values	Interactions
Labor Force Statistics from the Current Population Survey	Status	Total Labor Force (includes total armed forces) Armed forces Labor Force (includes resident armed forces) <other labor force variables>	Status is mostly labor force variables, but the concept of civilian/armed forces could be confused with industry (or even occupation).
	Class of worker	Wage and salary workers Private wage and salary workers Government wage and salary workers Self-employed workers Unpaid family workers Self-employed and unpaid family workers	Combination of occupation and industry sector values
	Industry	Private households (meaning private households in nonagricultural private wage and salary) Nonagriculture goods producing industries (mining, construction, and manufacturing) Service producing industries (nonag, code 040, except mining, construction and manufacturing) Nonagricultural industries (meaning: mining; constru; manufacturing; trans and public utilities; trade; fin, insurance, and real estate; sv industries including forestry and fisheries; and public admin) Construction Manufacturing Manufacturing, durables Manufacturing, nondurables Transportation, communication, and other public utilities (transportation and public utilities) Trade (wholesale and retail) Finance and service industries (meaning sv ind, 406, and tp26 ind codes 707-709) Service industries (includes forestry and fisheries and private households) Finance, insurance, and real estate Service industries except private households (miscellaneous	First set is what you get with status = labor force + military Second set (not explicitly listed is what you get with civilian labor force for status. Values include SIC codes.

		services on tabs, E.g., 55)	
		All nonagricultural industries except private households (this code has also been used in conjunction with 5:1 and 19:2 to identify other nonagriculture private wage and salary than private households, e.g., tabs 15) Long list of more specific industries	
Nonfarm Payroll Statistics from the Current Employment Statistics (National)	Industry	About 19 screens of codes SIC codes from 2-4 digits (division and lower) Variety of combinations: Total non-farm Total private Goods-producing Wholesale and retail trade Private nonagricultural Manufacturing, major manufacturing Durable goods Nondurable goods Service producing industry Private service-producing industry	Actual selection available depends on choice made for data type (unit of measurement, e.g., all employees, average weekly earnings, diffusion indexes) Note title, however – "Non-farm"
Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)	Detail	Total non-agricultural 1-digit industry (Industry division) 2-digit industry 3-digit industry A combination of industry divisions A combination of industries A residual cell	
	Industry	1-digit: all divisions 2-digit: subdivisions 3-digit: specific lists combination of industry divisions: goods producing service producing construction and mining services and mining (has asterisk) Combination of industries: Durable goods	Values depend on choice selected for detail. What do the asterisks mean? Note use of SIC codes with no explanation.

		Lumber & wood products including furniture Prim metal inds & fabr. Metal prods 33-36 mach & electronic & electric equip. nondurable goods textile mill products & apparel broadwoven and narrow fabric mills men's & boys' suits, coats, & furngs. Paper & printing & publishing Pulp, paper and paperboard mills 261, 262, 263, 266 chemicals & petroleum & coal products communications and public utilities wholesale trade retail trade A residual cell *R Other durable goods Other nondurable goods Other federal government Other state and local government	
Covered Employment and Wages	Industry	Division 2-digit SIC 3-digit SIC 4-digit SIC Nonclassifiable establishments All industries	
	ownership	Total covered Federal government State government Local government International government private	
	Establishment size	all	At the moment, there is no breakdown by size, but this is in the works.
Local Area Unemployment Statistics	none		

Geographic Profile	none		

Mass Layoff Statistics	IRC Code	Total all industries
		    Total all industry, private
		        Agriculture
		        Nonagriculture
		            Manufacturing
		                Durable goods
		                (list of specific durable goods industries, including "miscellaneous manufacturing industries")
		                Nondurable goods
		                (list of specific nondurable goods industries)
		            Non manufacturing
		            (list of divisions)
		        Not identified
		Government
		    Federal
		    State
		    Local
		
		Total, all reasons
		    Automation, bankruptcy, business ownership change, contract cancellation, contract completion, domestic relocation, energy-related, environment-related, financial difficulty, import competition, labor dispute, material shortage, model changeover, natural disaster, overseas relocation, plant or machine repair, product line discontinued, reorganization within company, seasonal work, slack work, vacation period, weather-related, other reasons, reason not reported
			Stands for "Industry, Reason, or Characteristic"
			Industry choices given first.
			Reasons are reasons for layoffs. Somewhat industry related.

		Characteristics = black, hispanic, women	
National Longitudinal Surveys (youth)	none		
Average Price Data	none		
Consumer Price Index – All Urban Consumers	none		
Consumer Price Index – Urban Wage Earners and Clerical Workers	none		
Consumer Price Index – All Urban Consumers (Old Series)	none		
Consumer Price Index – Urban Wage Earners and Clerical Workers	none		
Department Store Inventory Price Index	none		
Producer Price Index Revision – Current Series	Industry	2-digit SIC Codes 3-digit SIC Codes 4-digit SIC Codes <Non-SIC values> Crude processors Crude processors Crude processors Finished processors Primary processors Primary processors Semifinished processors Non-residential buildings Inputs to construction industries Final demand Final demand Finished processors Capital investment Other heavy construction	

		Highway and street construction Non-residential Maintenance and repair construction Residential New construction Consumers Consumers Primary processors Single-unit residential Multi-unit residential Semifinished processors Water and sewer supply construction Total manufacturing industries Total mining industries	
	Product	Long list of product codes, organized by industry code (including non-SIC) categories. Includes grouping such as durables, nondurables, goods less food and energy, etc.	Choice of industry variable controls list of products.
Producer Price Index Revision-Discontinued Series	Industry	Shorter list than above, mostly 4-digit, the only non-SIC is WINE	
	Product	Again, a shorter list than above.	Choice of industry variable controls list of products seen
Producer Price Index-Commodities	Group	All commodities Farm products Processed foods and feeds Textile products and apparel Hides, skins, leather, and related produ Fuels and related products and power Chemicals and allied products Rubber and plastic products Lumber and wood products Pulp, paper, and allied products Metals and metal products Machinery and equipment Furniture and household durables Nonmetallic mineral products Transportation equipment	These are commodities, which are loosely related to industry.

		Miscellaneous products Durability of product Industrial Commodities less fuels Industrial Commodities Farm products, processed foods and feeds Regional Refined Petroleum Special indexes Stage of processing	
	Item	Specific commodities within the groups in group variable. Includes the groups as values, as well.	Choice of group controls list of items seen.
International Price Index	Series name	Extensive list of commodities, products, manufactured items, agricultural items, etc.	Choice of index controls list of series. Does this list correspond to/harmonize with the commodities in PPI?
Consumer Expenditure Survey	none		
National Compensation survey	none		
Collective Bargaining Statistics – Historical (Private Sector)	Industry	All industries Manufacturing Nonmanufacturing Construction Goods-producing Service-producing	
Collective Bargaining Statistics – Historical (State and Local Government)	Sector	State and local government State government Local government	
Work Stoppage Data	none		
Employee Benefits Survey	Type	Medium and large private sector State and local governments Small private sector establishments	
Employment Cost Trends (Employment Cost Index)	Group	Wholesale and retail trade; excluding sales occupations Finance, insurance, and real estate; excluding sales occupations Wholesale trade; excluding sales occupations Goods-producing industries Goods-producing industries; excluding sales occupations	Values are combined with occupations; I've just listed those with industry components. Omitted values include just occupations, geography, union Values shown depend on choice for seasonal adjustment, compensation,

		Goods-producing industries; White-collar occupations	ownership.
		Goods-producing industries; White-collar occupations, excluding sales occupations	
		Goods-producing industries; Blue-collar occupations	
		Goods-producing industries; Service occupations	
		Service-producing industries	
		Service-producing industries; excluding sales occupations	
		Service-producing industries; White-collar occupations	
		Service-producing industries; White-collar occupations, excluding sales occupations	
		Service-producing industries; Blue-collar occupations	
		Service-producing industries; Service occupations	
		Non-manufacturing industries	
		Non-manufacturing industries; White-collar occupations	
		Non-manufacturing industries; White-collar occupations, excluding sales occupations	
		Non-manufacturing industries; Blue-collar occupations	
		Non-manufacturing industries; Service occupations	
		Construction	
		Manufacturing	
		Manufacturing - durable goods	
		Manufacturing - non-durable goods	
		Manufacturing; White-collar occupations	
		Manufacturing; White-collar occupations, excluding sales occupations	
		Manufacturing; Blue-collar occupations	
		Manufacturing; Service occupations	
		Transportation and Public Utilities	
		Transportation	
		Public utilities	
		Communications	
		Electric, gas, and sanitary services	
		Wholesale and retail trade	
		Wholesale trade	
		Retail trade	
		General merchandise stores	
		Food stores	

| | | Finance, insurance, and real estate
Banking, savings and loan, and other credit agencies
Insurance carriers, agents, brokers, and service
Insurance, excluding sales occupations
Services industries
Schools
Elementary and secondary schools
Services industries, excluding schools
Health services
Hospitals
Nursing and personal care facilities
Business services
Educational services
Public administration
Colleges and universities
Union; Manufacturing; Blue-collar occupations
Union; Manufacturing
Union; Non-manufacturing industries
Union; Goods-producing industries
Union; Service-producing industries
Non-union; Manufacturing; Blue-collar occupations
Non-union; Manufacturing
Non-union; Non-manufacturing
Non-union; Goods-producing industries
Non-union; Service-producing industries
Aircraft and parts manufacturing (SIC 372)
Aircraft and parts manufacturing (SIC 372); White-collar occupations
Aircraft and parts manufacturing (SIC 372);Professional, specialty, and technical occupations
Aircraft and parts manufacturing (SIC 372); Executive, administrative, and managerial occupations
Aircraft and parts manufacturing (SIC 372);Administrative support, including clerical, occupations
Aircraft and parts manufacturing (SIC 372); Blue-collar occupations
Aircraft and parts manufacturing (SIC 372); Precision, | |

		production, craft, and repair occupations Aircraft and parts manufacturing (SIC 372); Machine operators, assemblers, and inspectors occupations Aircraft manufacturing (SIC 3721) Aircraft manufacturing (SIC 3721); White-collar occupations Aircraft manufacturing (SIC 3721); Blue-collar occupations Aircraft engines and engine parts (SIC 3724) Aircraft engines and engine parts (SIC 3724); White-collar occupations Aircraft engines and engine parts (SIC 3724); Blue-collar occupations Aircraft parts and equipment, NEC (SIC 3728) Aircraft parts and equipment, NEC (SIC 3728); White-collar occupations Aircraft parts and equipment, NEC (SIC 3728); Blue-collar occupations Guided missiles and space vehicles manufacturing (SIC 3761) Guided missiles and space vehicles manufacturing (SIC 3761); White-collar occupations Guided missiles and space vehicles manufacturing (SIC 3761); Blue-collar occupations	
	Ownership	Civilian Private industry State and local government	
Nonfatal cases involving days away from work: selected characteristics	Category	Total private industry Industry division – agriculture Industry division – mining Industry division – construction Industry division – manufacturing Industry division – transportation and public utilities Industry division – wholesale trade Industry division -- retail trade Industry division – finance, insurance, and real estate Industry division -- services	Also includes age groups, days away from work, event or exposure, gender, length of service, nature of injury, occupation, part of body affected, race, source of injury.
	Industry	Industry divisions plus specific industries (3-digit?	

	Case type	Industry division or selected characteristic by detailed nature of condition Industry division or selected characteristic by detailed part of body affected Industry division or selected characteristic by detailed source of injury/illness Industry division or selected characteristic by detailed event or exposure Industry division or selected characteristic by detailed occupation Selected injury/illness characteristic by detailed industry Industry division or length of absence by age group Industry division or length of absence by gender Industry division or length of absence by race Industry division or length of absence by length of service	This describes how the data is broken out. Detailed industries are only available if appropriate case type is chosen.
Census of Fatal Occupational Injuries	Case type	Fatalities by detailed industry (all sectors) Fatalities by detailed private industry Fatalities by detailed government industry	Also has event, occupation, source of injury values.
	Industry	All workers Industry divisions 2- and 3-digit industries	
	Category	All industries All private industry Private industry division – agriculture Private industry division – mining Private industry division – construction Private industry division – manufacturing Private industry division – transportation and public utilities Private industry division – wholesale trade Private industry division – retail trade Private industry division – finance, insurance, and real estate Private industry division – services All government Federal government State government Local government	If case type is value without "detailed industry", Industry choices are presented here. Also includes age, gender, occupation, worker location, worker activity, primary and secondary source of injury
Occupational injuries and	Division	Private industry	Choice of division controls choices of

illnesses: industry data (pre-1989)		Agriculture, forestry, and fishing Mining Construction Manufacturing Transportation and public utilities Wholesale and retail trade Wholesale trade Retail trade Finance, insurance, and real estate Services	industry; only those within chosen division are shown.
	Industry	2-,3-,4(?) digit SIC codes within divisions	
Occupational injuries and illnesses: industry data (1989-current)	Division	Private industry Agriculture, forestry, and fishing Mining Construction Manufacturing Transportation and public utilities Wholesale and retail trade Wholesale trade Retail trade Finance, insurance, and real estate Services	Choice of division controls choices of industry; only those within chosen division are shown.
	Industry	2-,3-,4(?) digit SIC codes within divisions	
Major Sector Productivity and Costs Index	Sector	Manufacturing Manufacturing, Durable Goods Manufacturing, Nondurable Goods Business Nonfarm Business Farm Business Nonfinancial Corporations	Notice that sector is used with a different meaning here; not public/private.

Major Sector Multifactor Productivity Index	Sector	Manufacturing
		Nondurable goods (SIC 20-23, 26-31)
		    Food & kindred prod. (SIC 20)
		    Textile mills prod (SIC 22)
		    Apparel & related prod. (SIC 23)
		    Paper & allied prod. (SIC 26)
		    Printing & publishing (SIC 27)
			Notice that sector is used with a different meaning here; not public/private

		Chem. & allied prod (SIC 2) Petroleum refining (SIC 29) Rubber & plastic prod. (SIC 30) Durable good (SIC 24-25, 32-39) Lumber & wood prod (SIC 24) Furniture & fixtures (SIC 25) Stone, clay & glass (SIC 32) Primary metal ind. (SIC 33) Fabricated metal prod. (SIC 34) Ind. Machinery, comp. eq. (SIC 35) Electric & electr. eq. (SIC 36) Transportation equip. (SIC 37) Instruments (SIC 38) Misc. manufacturing (SIC 39) Elect & gas utilities (SIC 49) Private business Private nonfarm business	
International Labor Statistics	group	MANUFACTURING OUTPUT INDEX MFG AGGREGATE HOURS INDEX MFG AGGREGATE COMPENSATION INDEX MFG OUTPUT PER HOUR INDEX MFG HR COMP INDEX, NATL CURRENCY BASIS MFG HR COMP INDEX, US DOLLAR BASIS MFG UNIT LABOR COST INDEX, NATL CURRENCY BASIS MFG UNIT LABOR COST INDEX, US DOLLAR BASIS MANUFACTURING EMPLOYMENT INDEX MFG AVERAGE HOURS INDEX	Some indexes related to manufacturing. Others (not listed here) related to CPI, employment, etc.

II. Occupation and Related Variables, Selective Access

Series	Variable Name	Values	Interactions

Labor Force Statistics from the Current Population Survey	Occupation	Major Occupational Groups from OCS
		    Executive, administrative, and managerial
		    Professional specialty
		    Technicians and related support
		    Sales
		    Administrative support, including clerical
		    Service occupations
		    Precision production, craft, and repair
		    Machine operators, assemblers, and inspectors
		    Transportation and material moving
		    Farming, forestry, and fishing
		Several subgroups under each group.
		Aggregated total
		All nonfarm occupations
			Only available if all other variables are null.

	Class of worker	Wage and salary Self-employed Unpaid family worker	Interaction with industry: government, private, non-agricultural
Nonfarm Payroll Statistics from the Current Employment Statistics (National)	Datatype	Production workers	Limits industry choices to total private, Major divisions, some specific manufacturing
Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)	datatype	Production worker employment	Only available not seasonally adjusted.
Covered Employment & Wages	none		
Local Area Unemployment Statistics	none		

Geographic Profile <where does this actually live?>	Character code	Major Occupational Groups from OCS
		    executive, administrative, and managerial;
		    professional specialty;
		    technicians and related support;
			Major occupational group combined with:
			    age (16+)
			    Hispanic origin and age (16+)

		sales; administrative support, including clerical; service occupations; precision production, craft, and repair; machine operators, assemblers, and inspectors; transportation and material moving; farming, forestry, and fishing	Black/White and age (16+) Men/Women and age (16+)
Mass Layoff Statistics	none		
National Longitudinal Surveys (youth)	none		
Average Price Data	none		
Consumer Price Index - All Urban Consumers	none		
Consumer Price Index - Urban Wage Earners and Clerical Workers	Universe of queries	Wage earners Clerical workers	By definition of data collection: no choices.
Consumer Price Index - All Urban Consumers (Old Series)	None		
Consumer Price Index - Urban Wage Earners and Clerical Workers (Old Series)	Universe of queries	Wage earners Clerical workers	By definition of data collection: no choices
Department Store Inventory Price Index	none		
Producer Price Index Revision - Current Series	none		
Producer Price Index Revision - Discontinued Series	none		
Producer Price Index Revision - Commodities	none		
International Price Index	none		
Consumer Expenditure Survey	Column	Self-employed workers Total wage and salary earners Wage and salary earners: managers and professionals	Must choose "occupation of reference person" table

		Wage & salary: tech., sales, and clerical workers Wage & salary earners; service workers Wage & salary: construction workers & mechanics Wage & salary: operators, fabricators & laborers Retired Occupation: all other, including not reporting	

National Compensation Survey	Occupation	All occupations
		All occupations, excluding sales
		White collar occupations
		    White collar occupations, excluding sales
		        Professional specialty and technical occupations; subgroups
		        Engineers, architects, and surveyors; subgroups
		        Mathematical and computer scientists; subgroups
		        Natural scientists; subgroups
		        Health related occupations; subgroups
		        Teachers, college and university; subgroups
		        Teachers, except college and universities; subgroups
		        Librarians, archivists and curators; subgroups
		        Social scientists and urban planners; subgroups
		        Social, recreation, and religious workers; subgroups
		        Lawyers and judges; subgroups
		        Writers, authors, entertainers, athletes and professionals; subgroups
		        Technical occupations; subgroups
		        Executive, administrative, and managerial occupations
		        Executives, administrators and managers; subgroups
		        Management related occupations; subgroups
		        Sales occupations; subgroups
		        Administrative support occupations, including clerical; subgroups
		    Blue collar occupations
		        Precisions production, craft, and repair occupations; subgroups
		        Machine operators, assemblers, and inspectors; subgroups
		        Transportation and material moving occupations;	

		subgroups Handlers, equipment cleaners, helpers, and laborers; subgroups Service occupations, except private household Protective service occupations; subgroups Food service occupations; subgroups Health service occupations; subgroups Cleaning and building service occupations; subgroups Personal service occupations; subgroups	
	Level	Overall occupation average (no work level) Level 01 – 15 Not able to be leveled.	Must choose value for occupation that allows comparison with gov. job levels.
Collective Bargaining Statistics- Historical – Private Sector	none		
Collective Bargaining Statistics- Historical – State and Local Government	none		
Work Stoppage Data	none		
Employee Benefits Survey	none		
Employment Cost Index – Seasonal (Employment Cost Trends)	group	All workers Production and non-supervisory occupations All workers, excluding sales occupations White-collar occupations, excluding sales occupations White-collar occupations Executive, administrative, and managerial occupations Professional, specialty, and technical occupations Sales occupations Administrative support, including clerical, occupations Blue-collar occupations Precision, production, craft, and repair occupations Machine operators, assemblers, and inspectors occupations Transportation and material moving occupations Handlers, equipment cleaners, helpers, and laborers occupations Service occupations All workers, excluding sales occupations White-collar occupations, excluding sales occupations	Occupation values are combined with industry values in some instances. Also combined with union/non-union status. Note the frequent use of white-collar/blue-collar distinction A couple of duplicated values? This variable also contains some non-occupation industry values. Seasonally adjusted, Not seasonally adjusted, private industry, get all values

		Wholesale and retail trade; excluding sales occupations Finance, insurance, and real estate; excluding sales occupations Wholesale trade; excluding sales occupations Goods-producing industries; excluding sales occupations Goods-producing industries; White-collar occupations Goods-producing industries; White-collar occupations, excluding sales occupations Goods-producing industries; Blue-collar occupations Goods-producing industries; Service occupations Service-producing industries; excluding sales occupations Service-producing industries; White-collar occupations Service-producing industries; White-collar occupations, excluding sales occupations Service-producing industries; Blue-collar occupations Service-producing industries; Service occupations Non-manufacturing industries; White-collar occupations Non-manufacturing industries; White-collar occupations, excluding sales occupations Non-manufacturing industries; Blue-collar occupations Non-manufacturing industries; Service occupations Manufacturing; White-collar occupations Manufacturing; White-collar occupations, excluding sales occupations Manufacturing; Blue-collar occupations Manufacturing; Service occupations Insurance, excluding sales occupations Union; Blue-collar occupations Union; Manufacturing; Blue-collar occupations Non-union; Blue-collar occupations Non-union; Manufacturing; Blue-collar occupations Aircraft and parts manufacturing (SIC 372); White-collar occupations Aircraft and parts manufacturing (SIC 372);Professional, specialty, and technical occupations Aircraft and parts manufacturing (SIC 372); Executive, administrative, and managerial occupations Aircraft and parts manufacturing (SIC 372);Administrative	

		support, including clerical, occupations Aircraft and parts manufacturing (SIC 372); Blue-collar occupations Aircraft and parts manufacturing (SIC 372); Precision, production, craft, and repair occupations Aircraft and parts manufacturing (SIC 372); Machine operators, assemblers, and inspectors occupations Aircraft manufacturing (SIC 3721); White-collar occupations Aircraft manufacturing (SIC 3721); Blue-collar occupations Aircraft engines and engine parts (SIC 3724); White-collar occupations Aircraft engines and engine parts (SIC 3724); Blue-collar occupations Aircraft parts and equipment, NEC (SIC 3728); White-collar occupations Aircraft parts and equipment, NEC (SIC 3728); Blue-collar occupations Guided missiles and space vehicles manufacturing (SIC 3761); White-collar occupations Guided missiles and space vehicles manufacturing (SIC 3761); Blue-collar occupations .	
Nonfatal cases involving days away from work: selected characteristics	Category	Executive and professional specialty Technical, sales and administrative support Service Farming Precision production Operators, fabricators	Category also contains variables for race, sex, nature and source of injury, industry. (subtle difference between farming as occupation and industry, from non-expert's view.) Note 2 variables to choose occupation.
	Occupation	Detailed list as in NCS, above Includes military and last worked 1984 or before, and unclassifiable.	Available with # cases or median days lost chosen for datatype.
	Industry	All workers	This seems to contain both industry and product values.
Census of Fatal Occupational Injuries	Case type	Fatalities by detailed occupation	When this is selected, then the category variables change to include industry, age, etc., eliminating the general

			occupation listed below.
	Category	Executive and professional specialty Technical, sales and administrative support Sales Service Protective service Farming Precision production Construction Operators, fabricators Material movement Handlers and laborers Military	These values are available only if "detailed occupation" is not selected in case type. Note more detailed breakdown of occupations in which fatalities are more prevalent
	Industry	Starts out with "all workers"	I suspect by this they mean all industries, but this could be just a little confusing.
	Occupation	Long list of fine-grained occupations.	This is available only if "detailed occupation" is selected in case type.

Appendix C

I. Industry
II. Occupation

Background information mapping end user words and phrases associated with the *industry* and *occupation* concepts to BLS surveys/series

I. Industry-Related Concept/Term Groups from LSC

This appendix maps industry-related concept/term groups from LSC to series/variables. It also notes any related variables or constraints/dependencies between variables. Finally, it makes suggestions for additional kinds of clarifications that could be offered the user (in addition to synonym/near synonym guidance).

LSC tables: industry description, economy, producer costs, employer
NOTE: This file does not contain terms from industry type. We can assume that these must map in some way to SIC divisions, and detailed codes.

Concept: Industry

Terms: industry, business, firm, commerce, company, conglomerate, corporation, corporate entity, employer, enterprise, establishment, organization,
Common clarifications: Do you want to browse SIC? Do you want data at the industry division level, or a more specific industry? (give examples of division, 2-, 3-, 4-digit SIC industry
Series and variables:

Labor Force Statistics from the Current Population Survey
  Variables: Industry (Divisions, Combinations of Divisions, 3-digit SIC industries, private households.)
  Related variables: Status (labor force values)
   Class of worker (sector, occupation values)
  Constraints: If user wants 3-digit SICs, then Status = Civilian labor force, other variables (sex, race, ethnicity, age) = null.
  Clarification dialogue: Do you want data at the industry division level, or a more specific industry? <examples> Do you want to browse SIC?

Nonfarm Payroll Statistics from the Current Employment Statistics (National)
  Variables: Industry (Divisions, Combinations of Divisions, 2-4 digit SIC industries)
  Related variables: none
  Constraints: Data type selection controls industry detail level (and vice versa)
  Clarification dialogues or suggestions: Do you want data at the industry division level, or a more specific industry? <examples> Do you want to browse SIC?

Nonfarm Payroll Statistics from the Current Employment Statistics (State & Area)
  Variables: Detail (total non-agricultural, 1-digit industry (division), 2-ditig industry, 3-digit industry, combination of divisions, combination of industries, residual cell)
   Industry: All divisions, subdivisions, 3-digit industries, combinations, residual values
  Related variables: none
  Constraints: Detail value narrows down industry values available
  Clarification: need explanation of *residual*.
Do you want data at the industry division level, or a more specific industry? <examples> Do you want to browse SIC?

Covered Employment and Wages
  Variables: industry (divisions, 2-4 digit SICs, nonclassifiable establishments, all industries)

Related variables: ownership (total covered, federal government, state government, local government, international government, private)

Establishment size (all)

Constraints:

Clarification: Do you want data at the industry division level, or a more specific industry? <examples> Do you want to browse SIC?

Mass Layoff Statistics

Variables: IRC Code (Divisions, subdivisions, division combinations, some specific industries)

Related variables: none

Constraints:

Clarification: Do you want data at the industry division level, or a more specific industry? <examples> Do you want to browse SIC?

Producer Price Index Revision – Current Series

Variables: Industry (division, subdivision 2-4 digit SIC codes, limited combinations industries, stage of processing)

Related variables: Product (product codes, which are 5-6 digit SIC-related, product groupings)

Constraints: Choice of industry variable controls product selections available

Clarification: Do you want data at the industry division level, or a more specific industry? <examples> Do you want to browse SIC? Do you want to browse the product code lists? Explanation of stage of processing.

Producer Price Index Revision – Discontinued series

Same as Current Series (see above)

Collective Bargaining Statistics – Historical (Private Sector)

Variables: Industry (all industries, manufacturing, nonmanufacturing, construction, goods-producing, service-producing)

Related variables: none

Constraints:

Clarification: note explicit restriction to private sector. Note restricted choices of industry values.

Collective Bargaining Statistics – Historical (State and Local Government)

Variables: Sector (state and local government, state government, local government)

Related variables: none

Constraints: industry can only be state and/or local government

Clarification: note explicit restriction on industry/sector.

Employee Benefits Survey

Variables: Type (medium and large private sector, state and local governments, small private sector establishments)

Related variables: none

Constraints: industry can only be state and local governments, or all other industries.

Clarification: note restriction on industry values.

Employment Cost Trends (Employment Cost index)

Variables: Group (divisions, combinations of divisions, selected subdivisions, selected detailed industries.)
Related variables: Ownership (civilian, private industry, state and local government)
Constraints: values shown for group depend on selections made for seasonal adjustment, compensation, ownership.
Clarification: Anything useful to say about the level of detail for selected industries? If user chooses division/subdivision level for which detailed industry is available, should they be asked if they want a more detailed value? <Note the problem of learning – a person could enter a very detailed choice (e.g., guided missiles and space vehicles manufacturing) and be successful, but enter a more general choice (e.g., real estate) and be unsuccessful. Or the combinations with occupations, union?

Nonfatal cases involving days away from work: selected characteristics
Variables: Category (divisions or total private industry. Also includes non-industry choices.
Industry (divisions plus 3-digit industries)
Related variables: Case type (several choices including industry division, or "selected injury/illness characteristic by detailed industry")
Constraints: 2-3-digit industries only available if case type = detailed industry
Clarification: The dependency between case type and industry is straightforward once it is displayed. If person is interested in industry, can the other choices for category be minimized somehow?

Census of Fatal Occupational Injuries
Variables: Industry (all workers, divisions, 2- and 3-digit industries)
Category (all industries, all private industry, private industry divisions, various government choices. Also includes age, gender, occupation, worker, location, worker activity, primary and secondary source of injury.)
Related variables: Case type (by detailed industry (all sectors), by detailed private industry, by detailed government industry, also event, occupation, source of injury values).
Constraints: If case type value is other than one of the detailed industry choices, then category choices will include the division and sector values. If case type is one of the detailed industry choices, then industry choices include 2- and 3-digit industries.
Clarification. Dependency is a little tricky here.

Occupational injuries and illnesses: industry data (pre 1989)
Variables: Division (private industry, divisions)
Industry (2-4 digit SIC)
Related variables: none
Constraints: choice of division value dictates choices available for industry.
Clarification: Browse SIC?

Occupational injuries and illnesses: industry data (1989-current)
Same as pre-1989 above

Major Sector Productivity and Costs Index
Variables: Sector (manufacturing, durable goods, nondurable goods, business, nonfarm business, farm business, nonfinancial corporations)
Related variables: none
Constraints:

Clarification: Note different use of "sector". Explanation of limited number of choices? Explanation of meaning of "business"?

Major Sector Multifactor Productivity Index
      Variables: Sector (manufacturing, nondurable goods, durable goods, misc manufacturing, elect & gas utilities, private business, private nonfarm business, some 2-digit SIC within manufacturing)
      Related variables: none
      Constraints: none
      Clarification; note concentration on manufacturing, use of word "business".

International Labor Statistics
      Variables: Group (manufacturing, for various indexes, also cpi and employment indexes).
      Related variables: none
      Constraints:
      Clarification:

Concept: Division

Terms: division, industry division, employment sector, sector, business classification, industry segments.
Division values, division combinations.
Clarifications: See industry concept.
Can these terms be mapped to specific divisions or subdivisions?
      Regulated/deregulated businesses/industries
      e-commerce, dot-com (trade?)
      Hi-tech
      Internet companies (Communication?)
      Factories, plants (manufacturing?)
      Light industrial
      Seasonal businesses
Series and Variables: See industry concept.

Concept: Ownership

Terms: ownership, sector (ambiguous), private, , business, commercial, for-profit, corporate, privately-owned, privately held, public, government, federal, state, local.
Common clarification: If user query includes the term "sector", ask if they mean public/private distinction or industry division-level distinction. If public/private, continue with ownership. If division, move to industry concept.
For industries where public/private not specified, is correct assumption that both are included?
Can ownership also be extended to include "international" as possible value? If so, see international concept.
Series and Variables:

Labor Force Statistics from the Current Population Survey
      Variables: Class of worker (wage and salary workers, private wage and salary workers, government wage and salary workers, self-employed workers, unpaid family workers, self-employed and unpaid family workers)

Related variables: Industry (2 values seem pertinent, others are straight SIC) private households (meaning private households in nonagricultural private wage and salary, all nonagricultural industries except private households (this code has also been used in conjunction with 5:1 and 10:2 to identify other nonagriculture private wage and salary than private households, e.g., tabs 15))
Constraints: to get SIC division level, status = civilian.
Clarification: possible confusion on pertinent industry values?

Nonfarm Payroll Statistics from the Current Employment Statistics (National)
Variables Industry (pertinent values: total private, private nonagricultural, private service-producing industry. Other values are 2-4 digit SIC codes, plus division combinations.)
Related variables: none
Constraints: choices of values available for industry depends on choice for data type.
Clarification:

Nonfarm Payroll Statistics from the Current Employment Statistics (State and Area)
Variables: Industry (pertinent values: other federal government, other state and local government. Other values are divisions, 2-3 digit SICS, combinations.)
Related variables: Detail (pertinent values for industry only available if detail = a residual cell)
Constraints: see related variables
Clarification:

Covered Employment and Wages
Variables: ownership (total covered, federal government, state government, local government, international government private)
Related variables:
Constraints:
Clarification:

Mass Layoff Statistics
Variables: IRC Code (total all industries, total all industry, private, <lists of specific industries>, government, federal, state, local)
Related variables: none
Constraints:
Clarification: Assume that lists of specific industries is private only, or all?

Collective Bargaining Statistics – Historical (Private Sector)
Variables: industry (no ownership-related values)
Related variables: none
Constraints:
Clarification: Choice of this series makes private sector limitation.

Collective Bargaining Statistics – Historical (State and Local Government)
Variables: Sector (state and local government, state government, local government)
Related variables: none
Constraints:
Clarification: Choice of this series makes public sector/government limitation.

Employee Benefits Survey
Variables: Type (medium and large private sector, state and local governments, small private sector establishments)
Related variables: none
Constraints:
Clarification:

Employment Cost Trends (Employment Cost Index)
Variables: Ownership (civilian, private industry, state and local government)
Related Variables: Group (Divisions, industries, etc.)
Constraints: Values available for group depend on choice for ownership.
Clarification: Does "civilian" conflate labor force with ownerships?

Nonfatal Cases involving days away from work; selected characteristics
Variables: Category (total private industry, divisions, age groups, days, even/exposure, etc.)
Related variables: industry (specific industries).
Constraint:
Clarification: check assumption that divisions in category are all, not just private?

Census of Fatal Occupational Injuries
Variables: Case type (fatalities by detailed industry (all sectors), fatalities by detailed private industry, fatalities by detailed government industry. Also has event, occupation, source of injury values
Category (All industries, all private industry, private industry divisions, all government, federal government, state government, local government. Also includes age, gender, occupation, etc.
Related variables: Industry
Constraint: If case type is value without "detailed industry", industry choices are presented in category. If case type is value with 'detailed industry", industry variable presents divisions, 2-, 3-digit industries.
Clarification: Do industry values assume any ownership values?

Occupational injuries and illnesses: industry data (pre-1989)
Variables: Division (private industry, divisions)
Related variable: industry
Constraint: Values for industry depend on value chosen for division.
Clarification: Do industry values assume any ownership values?

Occupational injuries and illnesses: industry data (1989-current)
Same as pre-1989 above.

Major Sector Multifactor Productivity Index
Variables: Sector (HOM says this includes private only. In addition selected division and subdivision values, includes private business, private nonfarm business)
Related variables: none
Constraints:
Clarification: Although name of series doesn't specify, choice of this series limits to ownership = private.

Concept: Non-profit

Terms: non-profit, non-profit organization, not-for profit, third sector.
Clarification: Sometimes included, sometimes, not, in establishment surveys, see HOM. This seems sometimes to be at the discretion of the organization?
See ownership concept for loosely related variables.

Concept: Product, commodity

Terms: product, commodity, service-producing, goods-producing, durable goods, nondurable goods
Clarification: Would you like to browse the product specification? Note the difference between product, which continues SIC classification, and commodity, which doesn't. For those interested in goods-producing industries, service-producing industries, durable goods, nondurable goods, see industry.
Series and variables:

Producer Price Index Revision – Current Series
> Variables: Product (long list of product codes, organized by industry. Includes groupings such as durables, nondurables, goods less food and energy, etc.
> Related variables: Industry (2-4 digit SIC, stage of processing values)
> Constraint: Choice of industry limits list of products seen.
> Clarification:

Producer Price Index Revision – Discontinued Series
> See PPI current, above.

Producer Price Index-Commodities
> Variables: Group (all commodities, farm products, processed foods and feeds, textile products and apparel, hides, skins, leather, and related products, fuels and related products and power, chemicals and allied products, rubber and plastic products, lumber and wood products, pulp, paper, and allied products, metals and metal products, machinery and equipment, furniture and household durables, nonmetallic mineral products, transportation equipment, miscellaneous products, durability of product, industrial commodities less fuels, industrial commodities, farm products, processed foods and feeds, regional refined petroleum, special indexes, stage of processing)
> > Item (specific commodities with the group values)
> Related variables: none
> Constraint: choice of group controls list of items seen.
> Clarification: Commodities do not march directly with SIC; they cross industry categories. Note that some group values (e.g., stage of processing) do not expand to a list of specific commodities.

International Price Index
> Variables: Series name (extensive list of commodities, products, manufactured items, agricultural items, etc.)
> Related variables: none
> Constraint: Choice of index controls list of series.
> Clarification: Does this list correspond to/harmonize with the commodities in PPI?

Concept: Stage of processing

Terms: stage of processing, finished products, semifinished products, raw products, crude materials, raw materials, primary products

Clarification: Note the close relationship between processors (industries) and products. See those concepts for related terms, series, and variables.

Series and variables:

Producer Price Index Revision – Current Series

Variable: Industry (pertinent values: crude processors, finished processors, primary processors, semifinished processors, inputs to construction industries, final demand)

Related variable: Product (pertinent values: crude processors less agriculture, feed and energy, crude processors less energy, crude processors less agriculture, crude processors, crude agricultural processors, crude agricultural processors, crude agricultural and feed processors, crude mining processors, crude energy processors, crude feed processors, crude nondurables processors, crude durables processors, crude manufacturing processors less feeds, crude processors less agriculture and feeds, finished processors, finished mining processors, finished foods processors, finished nondurables processors, finished durables processors, finished manufacturing processors less foods, finished processors less food, primary agricultural processors, primary energy processors, primary processors less energy, primary processors less agriculture, primary processors, primary agricultural processors, primary agricultural and food processors, primary mining processors, primary food processors, primary nondurables processors, primary durables processors, primary manufacturing processors less foods, semifinished processors, semifinished agricultural processors, semifinished agricultural and food processors, semifinished food processors, semifinished nondurables processors, semifinished durables processors, semifinished manufacturing processors.)

Constraints: choice of industry constrains choice of product values shown)

Clarification: note that even though the variable is called product, and other values are products, these stage-of-processing values are processors.

Producer Price Index Revision – Discontinued Series

See PPI –current above.

Producer Price Index – Commodities

Variable: Group (Pertinent value: stage of processing)

Item (pertinent values: crude materials, crude foodstuffs and feedstuffs, crude nonfood materials except fuel, manufacturing, nonfood mat'ls exc fuel for manu excl sel items, construction, crude fuel, manufacturing industries, nonmanufacturing industries, crude materials less agricultural products, crude nonfood materials less energy, crude energy materials, crude materials less energy intermediate materials, supplies, and components, manufacturing, food manufacturing, nondurable manufacturing, durable manufacturing, components for manufacturing, materials and components for construction, processed fuels and lubricants, manufacturing industries, nonmanufacturing industries, containers, nonreturnable, supplies, manufacturing industries, nonmanufacturing industries, manufactured animal feeds, other supplies, intermediate materials less foods and feeds, intermediate foods and feeds, intermediate materials less food and energy, intermediate energy goods, intermediate materials less energy, finished goods, finished consumer goods, finished consumer foods, finished consumer foods, crude, finished consume foods, processed, consumer nondurable goods less food, consumer durable goods, capital equipment, manufacturing industries, nonmanufacturing industries, finished consumer goods excluding foods, finished goods, excluding foods, finished goods less food and energy, finished energy goods, finished goods less energy finished consumer goods less

food & energy, finished consumer goods less energy, consumer nondurable goods less food and energy, crude nonfood materials.

        Related Variable: none

        Constraint: Choice of group controls list of items seen

        Clarification: note apparent combination of industries and products in Item list? Note that apparent duplicated values have different codes.

International Price Index

        Variable: Series name (example pertinent values: other finished building materials, unfinished metals associated with durable goods, finished metals associated with durable goods, finished nonmetals)

        Related Variable Index (pertinent value, IL BEA End Use Import Indexes)

        Constraint: with this choice for Index, some values for series name seem to indicate stage of processing.

        Clarification: Do these indeed correspond to stage of processing? This would seem to be a small list, if so.

Concept: Establishment

Terms: establishment, company, business, firm, enterprise, corporation, employer, business site, workplace,

Common clarifications: Strictly speaking, this concept refers to a single unit of business. As such, information available concerns size. However, these terms are also used as near-synonyms for industry. See the industry concept for further information.

Concept: Establishment size

Terms: company size, corporation size, employer size, firm size, retail giant, small business, small-to-midsize companies, microenterprise, fast-growing business, major employers.

NOTE: many of these are really synonym for establishment. Note difference between industry size and establishment size. Also, think about different ways of considering size, e.g., number of employees, vs. product line, profits, and other financial measures. Does BLS only consider size in terms of numbers of employees?

In documentation on the Employee Benefits Survey, "small" means < 100 employees, medium and large are >= 100 employees.

Common clarifications: Need to indicate how size is measured (i.e., # employees). Also need to distinguish between industry size (which isn't included) and establishment size. See industry size concept.

Series and variables:

Covered Employment and Wages

        Variable: Establishment size (all)

        Related variable: none

        Constraint: none

        Clarification: At the moment, there is no breakdown by size, but this is in the works. For now, this is confusing in that it appears to offer a choice where there isn't one.

Employee Benefits Survey

        Variable: Type (medium and large private sector, state and local governments, small private sector establishments)

Related variable: none
Constraint: none
Clarification: Could clarify definition of small (< 100 employees) and medium/large (>= 100).

Concept: Industry size

Terms: industry size
NOTE important distinction between industry size and establishment size. See establishment size concept as potential confusion.
Common clarification: see establishment size.
Series and variables:
Labor Force Statistics from the CPS
Variable: Class of worker (Pertinent values: wage and salary workers, private wage and salary workers, government wage and salary workers) Yields number of workers within this class.
Related variables. Status (various labor force descriptors)
Industry (a variety of division, SIC values (see industry concept for details).
Constraint: status choice constrains list of industry values.
Clarification: This would be based on # employees per industry division or specific SIC code.

Nonfarm Payroll Statistics from the CES (National)
Variable: Data type: (pertinent value: all employees)
Related variables: industry (variety of divisions, combinations, 2-4 SIC codes)
Constraint: choice of data type influences selection available for industry. For all employees, get a reasonable sized selection, but not all.
Clarification: definition of industry size = # employees.

Nonfarm Payroll Statistics from the CES (State & Area)
Variable: data type (pertinent value: employment)
Related variables: detail (choice of level of industry description)
Industry (divisions, 2-digit industries, industry combinations)
Constraint: choice of detail controls available selections for industry.
Clarification: definition of industry size = # employees.

Concept: International business

Terms: foreign business, global business, international business, globalization (not precise).
Common clarification: may want to note limited information available from BLS, point to other locations? Suggest other countries' government sites? Many of these are not very close matches.
Series and variables:

Covered Employment and Wages
Variable: ownership (pertinent value: international government)
Related variable: none
Constraint: none
Clarification: government doesn't always = business in people's minds. This may not be a good match.

International Price Index
        Variable: None, really, just that this series reports international information.
        Related variable: none
        Constraint: None
        Clarification: This gives indexes/series of commodities, products, manufactured items, agricultural items, etc. Again, not necessarily a good match.

International Labor Statistics
        Variable: None, really, just that this series reports international information.
        Related variable: group (values indicate the particular index, some of which are specifically related to manufacturing)
        Constraint: none
        Clarification: This gives some specifically manufacturing-related information, so may be a slightly better match. Also has indexes related to CPI, employment, etc.

Concept: Unions

Terms: union, trade union, trade association, union shop, union-friendly business.
Clarification: BLS doesn't have a great deal of information about unions and union activity. Also point users to other sources.
Series and variables:

Mass Layoff Statistics
        Variable: IRC Code (pertinent value: labor dispute)
        Related variable: none
        Constraint: none
        Clarification: This is based on the questionable assumption that any labor dispute is related to union activity.

Collective Bargaining Statistics – historical (Private Sector)
        Variable: Series as a whole, which covers private sector union agreements.
        Related variable: none
        Constraint: none
        Clarification:

Collective Bargaining Statistics – Historical (State and Local Government)
        Variable: Series as a whole, which covers state and local government union/employee organization agreements.
        Related variable: none
        Constraint: none
        Clarification:

Work Stoppage Data
        Variable: Series as a whole, if one assumes that a work stoppage is often a union-organized effort.
        Related variable: none
        Constraint: none

Clarification: Not a strong match, since one would be making the questionable assumption that a work stoppage must involve union activity.

Employment Cost Index
        Variable: Group (pertinent values: union, non-union, also combined with various industry divisions and combinations.)
        Related variable: none
        Constraint: Values shown depend on choice for seasonal adjustment, compensation, ownership.
        Clarification:

Concept: Business age and life cycle

Terms: business age, business longevity, length in business, years in business, business success rates, business life, business start up, new corporations, new business formation, new business starts, start up, young companies, declining business, plant closings, bankruptcy, business failures, failure rate.
Common clarification: BLS doesn't collect much information here. See also Business development. Other places to refer users?
Series and variables:

Mass layoff statistics
        Variable IRC Code (pertinent values: bankruptcy,
        Related variables: none
        Constraint: none
        Clarification: see also business development concept.

Concept: Business development

Terms: business relocation, joint venture, bankroll, merger, takeover, acquisition, plant closing, bankruptcy, outsourcing, restructuring, seasonal businesses, seasonal economy, seasonal demand
Common clarification: BLS doesn't collect much information here. See also Business development. Other places to refer users?
Series and variables:

Mass Layoff Statistics
        Variable: IRC Code (pertinent values: automation, bankruptcy, business ownership change, contract cancellation, contract completion, domestic relocation, energy-related, environment-related, financial difficulty, import competition, labor dispute, material shortage, model changeover, natural disaster, overseas relocation, plant or machine repair, product line discontinued, reorganization within company, seasonal work, slack work, vacation period, weather-related, other reasons, reason not reported.)
        Related variables: none
        Constraint: none
        Clarification: se also business life cycle concept.

Concept: Regulation of industries

Terms: regulated business, deregulation, deregulated industries.

Clarification: No specific variable has this concept for values. However, presumably there is a list of commonly-recognized regulated or recently de-regulated industries (e.g., airline, banking, public utilities). Would it be worth suggesting some to users, or just referring them to the industry concept?
Series and variables: none.

Concept: Digital economy

Terms: digital economy, high-tech firms, high-tech industry, high-tech market, internet, internet business, I-business, net industry, new economy, virtual corporation, virtual office, technology incubator, dot-com (and variations) e-commerce, e-biz, e-tail, internet commerce
Clarification: As with regulated industries, no specific variable has this concept, but one could perhaps identify some industries that are related. Note especially the whole-sale/retail aspect. Would it be helpful to just steer users to the industry concept?
Series and variables: none.

Concept: Demographics of company owners

Terms: black owned business, minority business, minority owned business, family-owned business, mom-and-pop-stores, physician run companies, veteran and small business, women owned business, federally recognized native american business, native american business, indian companies.
Clarification: Does BLS collect any information here? If not, refer users to ??
Series and variables: none

II. Occupation-related concept/term groups from LSC

This appendix maps occupation-related concept/term groups from LSC to series/variables. It also notes any related variables or constraints/dependencies between variables. Finally, it makes suggestions for additional kinds of clarifications that could be offered the user (in addition to synonym/near synonym guidance).

LSC tables: employment, job categories, work arrangements, labor force, hours
NOTE: This file does not contain terms representing specific job types. We can assume that these must map in some way to SOC divisions or those found in OCS, and detailed codes. Also note that this deals with a very restricted notion of occupation. There are lots of ways of specifying information about occupation; here we're dealing just with occupation types, not things like shifts, job gain/loss, and so on.

Concept: Occupation

Terms: job, career, occupation, work, business <sometimes>, profession, trade, plus all terms referring to specific types of occupation.
Common clarifications: Need to give user list of occupation groups available (e.g., from SOC or OCS), so they can browse, or refer them to SOC to find appropriate group. See industry concept for similar treatment.
There is often the possibility of confusion with industry. E.g., the difference between farming (occupation) and agriculture(industry), or fishing, which is ambiguous.
Browsing opportunities from user's initial request to available values is important.
There may be some occupation terms that require special handling: farming/agriculture, businessman/woman/person.
Series and variables:

Labor Force Statistics from the Current Population Survey
       Variables: Occupation; (<major occupational groups from OCS> executive, administrative, and managerial; professional specialty; technicians and related support; sales; administrative support, including clerical; service occupations; precision production, craft, and repair; machine operators, assemblers, and inspectors; transportation and material moving; farming, forestry, and fishing; several subgroups under each group; aggregated total, all nonfarm occupations.
                Class of worker (wage and salary, self-employed, unpaid family worker.
      Related variables: none
      Constraints: only available if all other variables (age, industry, race, etc.) are null.
      Clarification dialogue:

Geographic Profile
      Variables: Character code (salient values.<major occupational groups from OCS> executive, administrative, and managerial; professional specialty; technicians and related support; sales; administrative support, including clerical; service occupations; precision production, craft, and repair; machine operators, assemblers, and inspectors; transportation and material moving; farming, forestry, and fishing)
      Related variables: none
      Constraints: occupation categories are combined with age, ethnicity, race, sex.

Consumer Expenditure Survey
    Variables: Column (salient values: self-employed workers, total wage and salary earners, wage and salary earners: managers and professionals; wage & salary: techn., sales, and clerical workers; wage & salary earners: service workers; wage & salary: construction workers & mechanics, wage & salary: operators, fabricators& laborers; retired; occupation: al other, including not reporting.
    Related variables: none
    Constraints: table value must = "occupation of reference person"

National Compensation Survey
    Variable: Occupation (All occupations, all occupations, excluding sales; white collar occupations; white collar occupations, excluding sales; professional specialty and technical occupations; <plus subgroups>; Engineers, architects, and surveyors <plus subgroups>; natural scientists <plus subgroups>; health related occupations <plus subgroups>; teachers, college and university <plus subgroups> teachers, except college and universities <plus subgroups>; librarians, archivists and curators <plus subgroups> social scientists and urban planners <plus subgroups> social, recreation, and religious workers <plus subgroups> lawyers and judges <plus subgroups> writers, authors, entertainers, athletes and professionals <plus subgroups> technical occupations <subgroups> executive, administrative, and managerial occupations, executives, administrators and managers <subgroups; management related occupations <subgroups> sales occupations <subgroups> administrative support occupations, including clerical <subgroups> Blue collar occupations; precision production, craft, and repair occupations <subgroups>; machine operators, assemblers, and inspectors <subgroups>; transportation and material moving occupations <subgroups>; handlers, equipment cleaners, helpers, and laborers <subgroups> service occupations, except private household; protective service occupations <subgroups> food service occupations ,subgroups>; health service occupations <subgroups>; cleaning and building service occupations <subgroups>; personal service occupations <subgroups>
    Related variables: level (overall occupation average (no work level), levels 01-15, not able to be leveled)
    Constraints: farming, fishing, forestry aren't included as possible values of occupation?
    Clarification: Occupation values seem to correspond with SOC groupings. Note the division into blue and white-collar occupations.

Employment Cost Index – Seasonal (Employment Cost Trends)
    Variable: group (salient values: all workers, production ad non-supervisory occupations, all workers, excluding sales occupations, white-collar occupations, excluding sales occupations, white-collar occupations, executive, administrative, and managerial occupations, professional, specialty, and technical occupations, sales occupations, administrative support, including clerical, occupations. Blue collar occupations, precision, production, craft, and repair occupations, machine operators, assemblers, and inspectors, handlers, equipment cleaners, helpers, and laborers, service occupations. These occupation values are also combined with industry groups, industries, union/non-union values)
    Related variable: none
    Constraints: more detailed occupation groups are available when they aren't combined with industry or union values.
    Clarification: These values are similar to OCS groups. Note again the use of blue and white collar.

Nonfatal cases involving days away from work: selected characteristics
    Variables: Category (executive and professional specialty; technical, sales and administrative support; service; farming; precision production; operators, fabricators.

Occupation: detailed list of values as in NCS above.
Related variables: Industry (all workers)
Constraints: category also contains variables for race, sex, nature and source of injury. Occupation choices are only available if datatype is # cases or median days lost.
Clarification: QUESTION: industry has the "all workers" value. Does that mean all industries, or all occupations? Does it really matter?

Census of Fatal Occupational Injuries
Variables: Case type (salient value: fatalities by detailed occupation.)
Category (Executive and professional specialty; technical, sales and administrative support; sales; service; protective service; farming; precision production; construction; operators, fabricators; material movement; handlers and laborers; military. Also other values for industry, age, etc.)
Occupation (Long list of fine-grained occupations. It isn't clear if it's based on SOC or not.)
Related variable: Industry (all workers)
Constraints: If case-type = detailed occupation, then you get to choose value for occupation, and the values for category exclude the occupations. Otherwise, you can choose occupation values for category, but have no choice for occupation.
Clarification: See above for industry comment.

Concept: White collar/Blue collar

Terms: white collar, white collar worker, non-manual worker, blue collar, blue collar workers, <low skill worker, unskilled worker, not exact synonyms>, pink collar <not used>
Common Clarifications: I'm kind of surprised that these terms are used; a search of the web site doesn't return anything. This is a good example of term collision. The National Compensation Survey actually uses these terms, and has SOC groups classified under them, which implies a technical definition. An end user may have different ideas about what falls into each class. As usual, browsing is necessary, and/or a list occupation classes that apply. The series listed below have an explicit value for white/blue: see also occupation.
Series and Variables:

National Compensation Survey
Variable: Occupation (salient values:, white collar occupations, white collar occupations, excluding sales; blue collar occupations
Related Variable: Level (overall occupation average (no work level), levels 01-15, Not able to be leveled.
Constraint: none
Clarification: white/blue are then subdivided into more specific classes.

Employment Cost Index – Seasonal (Employment Cost Trends)
Variable: group (salient values: white-collar occupations; white-collar occupations, excluding sales occupations; blue-collar occupations; goods-producing industries, these values then combined with industry groupings and divisions, unions, specific industries.

Concept: Self-Employed

Terms: self employed, self employment, freelance, independent contractor, consultant, independent consultant, unaffiliated worker

Common Clarifications: A consultant may instead work for a consulting firm. The series listed below have an explicit value for self-employed: see also occupation.
Series and Variables:

Labor Force Statistics from the Current Population Survey
        Variable: Class of worker: (wage and salary, self-employed, unpaid family worker.
        Related Variable: Occupation (lists major OCS groups)
        Constraints: Class of worker interacts with industry: government, private, non-agricultural
        Clarification: Many occupations may either work for a company, or be self-employed. But it's hard to think of a self-employed production worker.

Consumer Expenditure Survey
        Variable: Column (salient variable: self-employed workers)
        Related variable: Level (overall occupati9on average (no work level), levels 01-15, not able to be leveled.)
        Constraints: none
        Clarification:

Concept: Contingent worker

Terms: contingent worker, temp, temporary worker, temporary job, temp help, temp staffing, seasonal worker, contract worker
Common Clarifications:
Series and Variables: QUESTION: are these tracked by occupation, or by industry?

Concept: Government employees

Terms: civil service employees, federal employees, federal government employees, municipal employee, city officials, public employee, public servant, state worker
Common Clarifications: This is actually a confusion with industry. Government worker is a sector, not an occupation. So see industry concepts, as well as series below.
Series and Variables:

National Compensation Survey
        Variable: Level (overall occupation average (no work level); levels 01-15, not able to be leveled.)
        Related variable: Occupation (lots of values)
        Constraints:
        Clarification: This variable represents a correspondence with federal government job classification levels.

Concept: Military personnel

Terms: military personnel, armed forces personnel, specific branches of the military.
Common Clarifications: This may require special treatment, since many surveys exclude military from the sampling universe. Also, specific occupations in the military would come under the general civilian classification for many purposes (e.g., skills required from the OOC). Series below lists military specifically.
Series and Variables:

Nonfatal cases involving days away from work: selected characteristics
>Variable: Occupation: (salient value: military)
>Related Variable: Datatype (salient values # cases or median days lost)
>Constraint: Must choose one of the values given for datatype in order to see detailed list of occupations.
>Clarification:

Census of Fatal Occupational Injuries
>Variable: Category (salient value, Military)
>>Occupation (salient value, military occupations)
>Related variables: Case type (anything other than fatalities by detailed occupation)
>Constraint: If case-type = detailed occupation, then you get the detailed list of occupations, and can choose military in occupation. Otherwise, you can choose military in the category variable.
>Clarification:

Concept: Occupation definitions

Terms: list of occupation titles, job titles, occupation codes, occupational classification system, equal employment opportunity job codes, job descriptions, occupational characteristics, position descriptions.
Common Clarifications
Resources:
Note: these don't lead to series, but rather to other types of publications. I've just given URLs here.

Occupational Outlook Handbook
>http://www.bls.gov/ocohome.htm

Career Guide to Industries
>http://www.bls.gov/cghome.htm

Standard Occupational Classification (SOC) System
>http://www.bls.gov/soc/soc_home.htm

Occupational Compensation Survey Occupational Job Descriptions
>http://www.bls.gov/ocsjobde.htm
>NOTE: OCS is discontinued, but the data is available through 1997.

Concept: Gender and occupations

Terms: male occupations, female occupations, pink collar occupations, gender roles, traditionally female occupations
Common clarifications: As a user query, this concept is a little backwards. A user could look at an occupation or occupation group and get division by sex, and from there infer whether it was primarily a male or female occupation, whether the division had changed over time, etc. But the question "what are female occupations" cannot be answered directly. Similarly, a question about "traditionally female occupations" requires some data analysis and/or cultural judgments.
Series:
See occupation concept.

Concept: Dangerous occupations

Terms: dangerous occupations, dangerous jobs, dangerous work, hazardous duty, hazards, stressful jobs, stressful occupations, ergonomics <indirect>

Common clarifications: Although asking for "dangerous occupations" would require a judgment by the user on what constitutes a dangerous level of injury/death, a user could be steered toward the series that give injury/death data, and let them make the judgment.

Series:

Nonfatal cases involving days away from work: selected characteristics

       Variables: Category (pertinent values: Executive and professional specialty, Technical, sales and administrative support, Service, Farming, Precision production, Operators, fabricators. Also contains values for race, sex, nature and source of injury, industry)

               Occupation (Detailed list as in NCS (20) above.

       Related variables: Data type

       Constraints: Detailed occupation values only available if datatype = # cases or median days lost. Otherwise, limited to broad values in category.

       Clarification:

Census of Fatal Occupational Injuries

       Variables: Case type (pertinent value: Fatalities by detailed occupation)

              Category: (Executive and professional specialty, Technical, sales and administrative support, Sales, Service, Protective service, Farming, Precision production, Construction, Operators, fabricators, material movement, Handlers and laborers, Military)

              Occupation: (Long list of fine-grained occupations <SOC?>)

       Related Variables: none

       Constraints: If case type = "by detailed occupation", then category occupation values aren't available, but Occupation values are. Otherwise, only category values are available.