

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

UNECE Work Session on Statistical Dissemination and Communication
(12-14 September 2006, Washington D.C., United States of America)

Topic (iii) How to present metadata

CREATING A USER-FRIENDLY GLOSSARY AND OTHER METADATA MAKEOVERS

Supporting Paper

Submitted by Energy Information Administration, United States¹

I. INTRODUCTION

1. Creating helpful metadata on the web is often overlooked or the last thing to be done, and sometimes we hold on to the model we have used for years in the paper world. This paper discusses the importance and power of making transitions from paper metadata to electronic metadata: How using the capabilities of the web to provide metadata can enhance the user experience on your site.
2. EIA has had a busy year redesigning our website and improving our search engine. In addition, we have looked at other areas for improvement on our website and will share three experiences involving metadata.
3. First, in reorganizing a huge amount of energy consumption data, we ran into a balancing act of showing too much information about the data versus not showing enough. What is a good balance to help customers find what they are looking for?
4. Next, our old glossary was just an alphabetical list of words. Our new web-friendly glossary allows users to search differently for terms, allows staff to create custom glossaries, and provides on-the-spot popup definitions where needed.
5. Finally, an area of some frustration to many authors and users alike is the use of footnotes on the web. What is the best way to give definitions, what is the best placement for footnotes, what is a good format? We have been thinking about these questions.

II. TOO MUCH VS. TOO LITTLE ON THE PAGE

6. Strange as it may sound, sometimes having too much information (too many links, too much description, too many places to navigate) on a web page can actually hide the data customers are looking for. You can see the problem with a quick review of our old energy consumption page.

¹ Prepared by [Melinda Hobbs melinda.hobbs@eia.doe.gov].



7. Customers used to ask how they could get to the data and whether the data they did find are the most recent we have. There was very little information on this page telling people the kinds of data available. Users tended to look at the long and interesting titles on the right side bar, which were in fact supporting analyses, and would get lost very quickly from this page.

8. If they happened to get beyond this page, the next page gave a long, complete set of data tables available for the survey. Even though the title of the page clearly stated the data for 1998, because this survey is only fielded every four years (with up to 18 additional months for posting), the date looked very old. How would customers know this was, in fact, the most recent data? We were diligent about labeling the dates of the data, but not about explaining why it seemed so old. There was a link to information about the data, but how many people would take that side trip or even see the link with so much on the page?

9. Another problem was that when the new data did become available, this table was replaced with all new links, making it difficult for users to get back to the previous data. If you happened into these pages through a search and just got dropped in, you probably would not know what you were looking at.

10. A final problem was that each consumption area (residential, commercial, manufacturing) displayed their data in a different format. (They were different groups so they thought why not have a different look?). The theory behind the residential energy page, for example, was to have everything above the fold.

IMPORTANT NOTICE: The 2005 Residential Energy Consumption is now in the field. Personal interviews with selected housing units for the 2005 Residential Energy Consumption Survey (RECS) began in mid-August and will continue for approximately six months. The interviews are being conducted by the National Opinion Research Center at the University of Chicago.

RECS household characteristics data are anticipated to be available in the Fall of 2006. The consumption and expenditures data are anticipated to be available in the Spring of 2007. For more information about the 2005 RECS, contact Stephanie Battles, RECS Survey Manager at stephanie.battles@eia.doe.gov or call her at 202/586-7237.

The Residential Energy Consumption Survey (RECS) provides information on the use of energy in residential housing units in the United States. This information includes:

- the physical characteristics of the housing units,
- the appliances utilized including space heating and cooling equipment,
- demographic characteristics of the household,
- the types of fuels used, and
- other information that relates to energy use.

11. It's a good theory, but it has to be usable and understandable. All the actual information was hidden under the dropdowns and there was no way to tell which topic was more important. This design, meant to help the users, ended up being more confusing.

12. During the redesign we decided to combine all the consumption surveys on to one second-level page instead of the four we had previously. To highlight the data and allow the customers to get data faster, we pulled out similar kinds of information across all the consumption surveys. We aimed for consistency on each survey page.

Main consumption page summarizes the similarities between the surveys and leads customers down into more detailed data with consistent layouts.

13. On the new pages, it's much easier to see the kinds of data available, the years available, along with an explanation of the survey and more detailed information about the data or survey. Putting everything out on the website without some sort of organization hid the hierarchy of information and was a disservice to our customers.

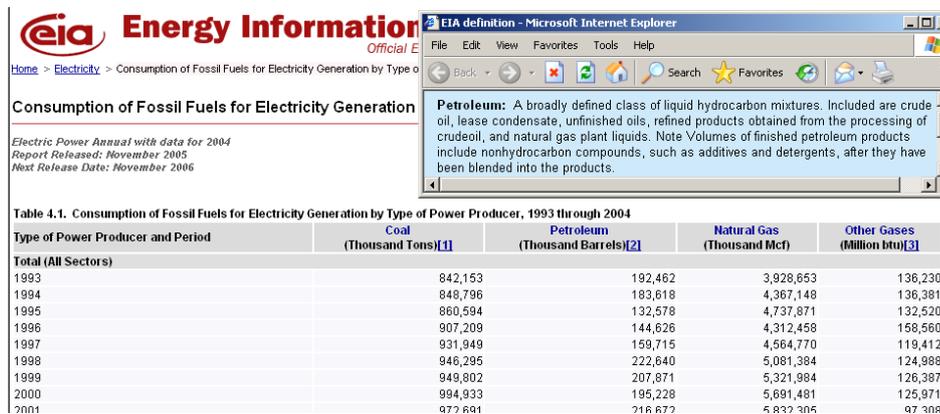
III. USEFUL GLOSSARY

14. Our old glossary was like many others, with pages of like-lettered definitions in alphabetical order. EIA's glossary began as many small, separate, product-specific glossaries. As a result of different authors and different offices, sometimes the definitions of the same term didn't match.

15. EIA undertook to consolidate all the terms and agree on common definitions. In our new glossary, all the terms mean the same thing. But each publication or report still wanted its own glossary so they manually had to pull the terms they needed and package them into a customized glossary.

16. Our new web glossary puts each of the terms into a database. This allows authors to pull out subsets of terms (for products that do not need the entire glossary) without having to "create their own." Customers can now search the glossary for a specific term, several terms, and related terms.

17. Because the glossary is now a database, the terms can be linked to from tables and publications so the definition can pop up with a click on the term in context.



The screenshot shows the EIA website page titled "Consumption of Fossil Fuels for Electricity Generation" with a table of data. A pop-up window from Microsoft Internet Explorer displays the definition for "Petroleum": "A broadly defined class of liquid hydrocarbon mixtures. Included are crude oil, lease condensate, unfinished oils, refined products obtained from the processing of crudeoil, and natural gas plant liquids. Note Volumes of finished petroleum products include nonhydrocarbon compounds, such as additives and detergents, after they have been blended into the products."

Type of Power Producer and Period	Coal (Thousand Tons)[1]	Petroleum (Thousand Barrels)[2]	Natural Gas (Thousand Mcf)	Other Gases (Million Btu)[3]
Total (All Sectors)				
1993	842,153	192,462	3,928,653	136,230
1994	848,796	183,618	4,367,148	136,381
1995	860,594	132,578	4,737,871	132,520
1996	907,209	144,626	4,312,458	158,560
1997	931,949	159,715	4,564,770	119,412
1998	946,295	222,640	5,081,384	124,888
1999	849,802	207,871	5,321,984	126,387
2000	894,933	195,228	5,691,481	125,971
2001	877,681	216,672	5,837,305	97,308

18. When this feature was shown to internal staff they actually applauded! Having the definition so readily available is very helpful to customers. This database is useful to staff internally for customizing specific glossaries and it's helpful to customers looking for definitions.

IV. NOW, WHAT ABOUT FOOTNOTES ON THE WEB?

19. Footnotes and sourcing isn't a glamorous issue so it doesn't get as much attention, at least in EIA. Most people seem to concentrate on the main information and not pay too much attention to the inclusion or format of the sourcing.

20. A more useful way to show footnotes is having them pop up like the new glossary definitions shown above or at least having them linked so users don't have to scroll up and down to see footnotes. Web authors need to move away from "thinking in paper" even though having notes all at the bottom or on a separate page is needed for printing.

21. Our data are still primarily being footnoted and sourced from within static tables, which forces interested users to have to search on their own for sources and definitions. Sometimes the footnotes at the bottom of the page are live links to the sources, while other times they are just typed links, depending on the legacy system that generated the table or whether the file was "hand made." And, even if the table were created new from the beginning, whether the footnotes were relevant, consistent, and live depends on the skill of the person who made it.

22. In many cases, the initial metadata is stored in a database but placing it as a live link where it is needed is limited by the type of data table it is attached to. Legacy dissemination formats have not been rethought and reworked because we have had to focus more on the data collection side.

23. A final problem is the inconsistency of footnotes for static and dynamic tables, publications, papers, charts and graphs. Here is a good example of footnoting.



Official Energy Statistics from the U.S. Government

[Glossary](#)

[Home](#) > [Electricity](#) > Existing Capacity by Energy Source

Existing Capacity by Energy Source

*Electric Power Annual with data for 2004
Report Released: November 2005
Next Release Date: November 2006*

Table 2.2. [xls](#) [pdf](#) format

Table 2.2. Existing Capacity by Energy Source, 2004 (Megawatts)

Energy Source	Number of Generators	Generator Nameplate Capacity	Net Summer Capacity	Net Winter Capacity
Coal ^[1]	1,526	335,243	313,020	315,364
Petroleum ^[2]	3,175	37,970	33,702	37,339
Natural Gas	3,048	256,627	224,257	241,391
Dual Fired	3,003	193,115	172,170	184,399
Other Gases ^[3]	119	2,535	2,296	2,259
Nuclear	104	105,560	99,628	101,377
Hydroelectric Conventional ^[4]	3,995	77,130	77,641	77,227
Other Renewables ^[5]	1,608	21,113	18,763	19,000
Pumped Storage	150	19,569	20,764	20,676
Other ^[6]	42	754	700	716
Total	16,770	1,049,615	962,942	999,749

[1] Anthracite, bituminous coal, subbituminous coal, lignite, waste coal, and synthetic coal.
 [2] Distillate fuel oil (all diesel and No. 1, No. 2, and No. 4 fuel oils), residual fuel oil (No. 5 and No. 6 fuel oils and bunker C fuel oil), jet fuel, kerosene, petroleum coke (converted to liquid petroleum, see Technical Notes for conversion methodology), and waste oil.
 [3] Blast furnace gas, propane gas, and other manufactured and waste gases derived from fossil fuels.
 [4] The net summer capacity and/or the net winter capacity may exceed nameplate capacity due to upgrades and overload capability of hydroelectric generators.
 [5] Wood, black liquor, other wood waste, municipal solid waste, landfill gas, sludge waste, tires, agriculture byproducts, other biomass, geothermal, solar thermal, photovoltaic energy, and wind.
 [6] Batteries, chemicals, hydrogen, pitch, purchased steam, sulfur, and miscellaneous technologies.
Notes: Where there is more than one energy source associated with a generator, the predominant energy source is reported here. Totals may not equal sum of components because of independent rounding.
Source: Energy Information Administration, Form EIA-860, "Annual Electric Generator Report."

V. CONCLUSION

24. Of course content is most important on our websites. But without useful and understandable and findable metadata, users won't be as successful with our information. Metadata shouldn't be a scary word; it's a group of tools to help users find their way through piles of sometimes confusing statistics. Especially important is having metadata using web capabilities to allow people to link, jump out and jump back in, and search. Metadata shouldn't be just adding information at the end like in an old-fashioned paper report or having it in static files.