

Railroad Data Audit
By Markham Heid

The database produced by the Federal Railroad Administration Office of Safety Administration is very comprehensive in the sheer amount of data it provides and many of the ways it offers users to search and download the data. It makes some of the more intriguing data, such as the casualty statistics, available for download in different formats. However, the organization of this downloadable data is cumbersome and its fields are occasionally unintelligible.

For example, when downloading casualty data for 2009, the explanations for casualties are not standardized and inexplicably span three separate columns that must be joined. Not sure why they would do this.

Also, the functionality for the FRA database is not as complete as it at first seems. Many of the search queries do not work properly, or at all, and so it's difficult to find much of the contextual information necessary to parse findings. For example, the number of miles each railroad travels by state is not available. So while casualty numbers and other per state statistics are available, it's very difficult to determine the significance of the numbers on anything other than a national basis.

Also, the FRA database treats all railroad lines equally. Amtrak, a primarily commercial passenger railroad, is treated exactly like Union Pacific or BASF, which are primarily cargo railways. When looking at data for hundreds of railroad lines, this makes it almost impossible to judge trends in casualties based on the type of people working or utilizing the various railroads.

The FAA likely discriminates between commercial and transport airline carriers, but the FRA makes no such discrimination. We're dealing here with completely different industries with vastly different types of people involved with the day to day functions of the railroads. The FRA needs to develop designations for types of railroad companies, and should include that information in its database fields.

The FRA is kind enough to provide phone numbers for specific individuals on its Web site, however it provides no information on the 11 contacts listed, so it's basically a crap shoot as to who handles what section of the data and would be most appropriate to contact. Also, no email is provided beyond a general "feedback" page.

Also, while the database provides many different and helpful search criteria for exploring the data, there is no simple way to extract one large data file with all of the disparate pieces of information. While this setup initially seems like a convenience, it's more helpful to someone trying to locate a single number, as opposed to a journalist hoping to explore in depth the FRA findings.

All in all, this is a very complete database. However, the exclusion of state-by-state operational data for each railroad, and the failure to establish operational designations for the various railroad types are major failings that limit the amount of conclusions journalists could draw without a lot of additional contact with other sources outside of the FRA.