

DRB Minutes for 11/27/06

1. SIPP/IRS/SSA puf (b) (6) gave an overview of the synthetic data project, in particular what data items are involved in each file. The disclosure model we are following for the puf is oriented toward the ability or inability to link to the existing SIPP public use file. This makes sense since a SIPP-like, file which pulls together a broad range of demographic and financial data, is the best that an attacker can hope to have. And the file itself is available and unlike other data sources offers the guarantee that the true “corresponders” are in there. Measures have been taken to make sure the synthesized components do not correspond too closely to the record components they replace, both at single points and longitudinally. Linkage to SIPP seems the most likely avenue of disclosure ... i.e., that the additional information from the SIPP/IRS/SSA puf might endanger the protection on the old file. (b) (6) went into some detail on the matching efforts ... they tried 3 approaches, none of which lead to a significant match rate. And all of which used knowledge an attacker would not have. The correctly linked records appear to be rare random occurrences and are indistinguishable from incorrect associations, which abound. Linking this file and the SIPP puf would not increase the overall risk due to the low quality of such a linkage even in circumstances favorable to the attacker. This release was approved.

2. More discussion of (b) (6) Still not done.
3. Handout on the addendum as implemented for CE.
4. USDA This goes back to 80 and 90 data. This is pumsable except of the urban/rural designation. This was approved and since it involves data prior to the rounding rules, it need not be rounded.
5. 2010 planning data. The proposed release is clearly pumsable and is approved. Guidance was sought over how detailed a geography would be allowed; a 50-50 rule was agreed to ... 50,000 pop and 50 unweighted cases for a given ethnicity/race.