

## Worksheet for estimating cost of data deposits

---

The NIH policy on data sharing acknowledges that applicants can request funding for data sharing or ask for funding if not included in funding awards.

*Applicants may request funds in their application for data sharing. If funds are being sought, the applicant should address the financial issues in the budget and budget justification sections. Some investigators have more experience than others in estimating costs associated with preparing the dataset and associated documentation, and providing support to data users. As investigators gain experience with the process, their ability to estimate costs will improve. Investigators working with archives can get help with data preparation and cost estimation. Investigators who are concerned about paying for data-sharing costs at the end of their grant can make prior arrangements with archives. Investigators facing considerable delays in the preparation of the final dataset for sharing should consult with the NIH program about how to manage this situation, such as requesting a no-cost extension.* (NIH Data Sharing Policy and Implementation Guidance; 5 March 2003)

### Part A: Cost per hour of effort

Median FTE salary of persons involved in preparing dataset. . . . . \$ \_\_\_\_\_

Median employee fringe benefits costs (\_\_\_% of salary). . . . . \$ \_\_\_\_\_

Supplies and equipment costs per FTE (\_\_\_% of salary). . . . . \$ \_\_\_\_\_

**Total direct cost (TDC).** . . . . . **\$ \_\_\_\_\_**

Facilities and administrative (F&A; \_\_\_% of TDC). . . . . \$ \_\_\_\_\_

**Total cost per median FTE of effort (TDC + F&A).** . . . . . **\$ \_\_\_\_\_**

Total work hours per yr (52wks x 40hrs). . . . . 2,080 hrs

Less vacations days (\_\_\_ days). . . . . - \_\_\_ hrs

Less holidays (\_\_\_ days). . . . . - \_\_\_ hrs

Less sick days and family leaves (\_\_\_ days). . . . . - \_\_\_ hrs

Less time for professional development (\_\_\_ days). . . . . - \_\_\_ hrs

Net work hours. . . . . \_\_\_ hrs

**Hourly rate (total cost per median FTE of effort ÷ net work hours).** . . . . . **\$ \_\_\_\_\_**

### Part B: Person hours of work

Staff meetings to devise the deposit plan and to identify variables to be re-coded, redacted, and collapsed (no. of meetings x median number in attendance). . . . . \_\_\_ hrs

Person hrs of programming time to re-code, redact, and collapse and to create the de-identified dataset. . . . . \_\_\_ hrs

Person hrs of time verifying dataset to ensure proper coding, stripping,

---

 Worksheet for estimating cost of data deposits

and collapsing.....	___ hrs
Person hrs preparing data dictionary and accompanying documentation. ....	___ hrs
Person hrs checking accuracy of data dictionary and accompanying documentation.. ....	___ hrs
Testing the data file for robustness of user instructions.....	___ hrs
Challenges to de-identification (to see if persons can be re-identified).....	___ hrs
Preparation and submission of coordinating center IRB submission regarding deposit. ....	___ hrs
Preparation of memo detailing deposit plan for distribution to center directors for submission to their respective IRBs. ....	___ hrs
Hours dealing with queries from coordinating center IRB regarding deposit. ....	___ hrs
Hours dealing with clinics concerning their IRB submissions regarding deposit. ...	___ hrs
Hours of effort collecting individual IRB deposit approvals.....	___ hrs
Revision of the dataset based on interaction with coordinating center IRB and other IRBs. ....	___ hrs
Interaction with funding agency and repository custodian regarding deposit. ....	___ hrs
Transmission of dataset to repository.....	___ hrs
Other efforts.....	___ hrs
<b>Total hours.</b> .....	<b>___ hrs</b>

*Part C: Difficulty multiplier*

- a. Number of data collection sites.....
- b. Number of data generation sites (reading centers, specimen laboratories, etc). ..
- c. Number of IRBs (typically equivalent to the number of study centers). ....
- d. **Sum** (items a thru c). ....
- e. Difficulty multiplier
 

<b>Sum</b> ≤ 20 (multiplier 1.00). ....	( )
<b>Sum</b> > 20 but < 40 (multiplier 1.05). ....	( )

---

 Worksheet for estimating cost of data deposits
 

---

**Sum**  $\geq$  40 (multiplier 1.10)..... ( )

*Part D: Estimated total cost for deposit*

Total hours from Part B. .... hrs

Difficulty multiplier from Part C. ....

**Total cost (Total hours x hourly rate x difficulty multiplier)..... \$\_\_\_\_\_**

Note 1: The median salary cost method of calculation assumes the mix of effort involved is consistent with the method. If that assumption is markedly false adjust the hourly calculation to reflect reality.

Note 2: The calculations do not include efforts in answering questions from users of the datasets.