



R²

Towards Fully Replicable Data Analysis in an Increasingly Connected World

Florio O. Arguillas, Jr., Ph.D.

William C. Block, Ph.D.

Jeremy Williams

CISER

2016 IASSIST Conference

June 1-3

Scandic Hotel

Bergen, Norway



International Association for Social Science
Information Services & Technology

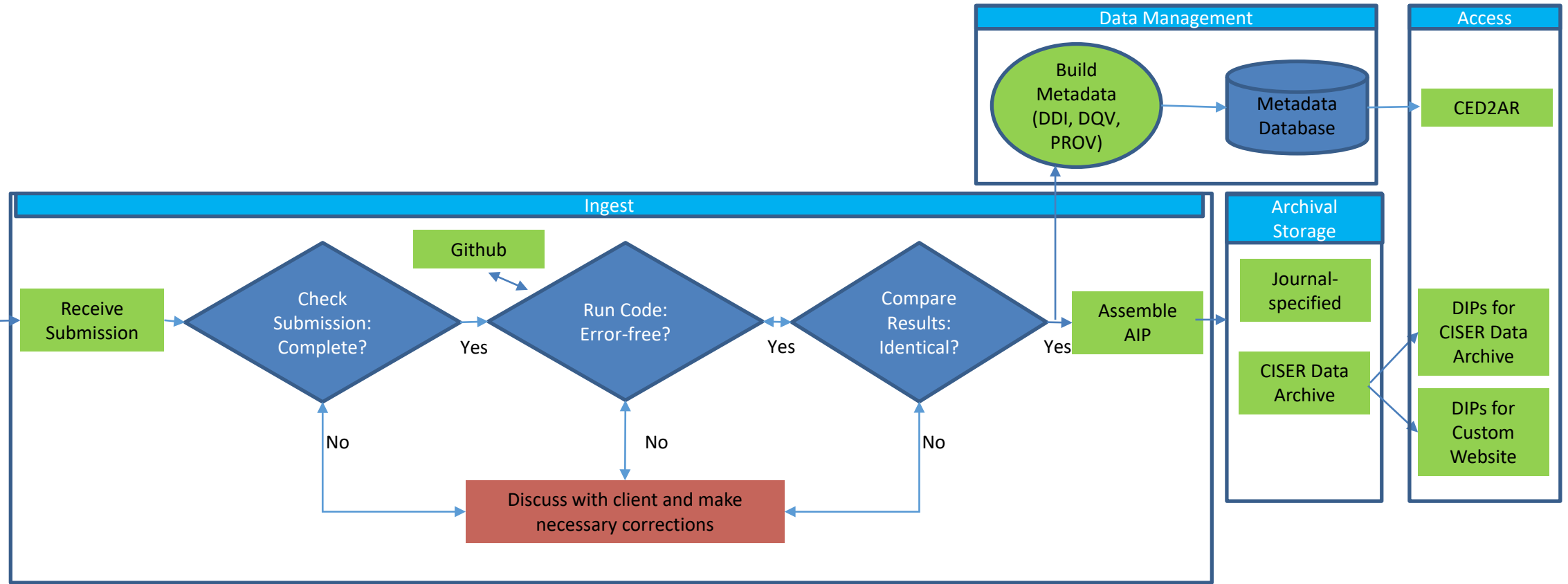


Step 1

You've got Mail!

Step 2

Client Deposits





| 1995 | 1995 | 1999 | 1999 |
|--------|--------------|--------------|--------------|
| Wanted | Consistent | Inconsistent | Inconsistent |
| Wanted | Inconsistent | Consistent | Inconsistent |
| Wanted | Inconsistent | Inconsistent | Consistent |

FIG. 1—First Dependent Variable Construction

TABLE 1a
Consistency of Women's Wanted/Not Wanted Status Between Time 1 and Time 2 About 18 Months Postpartum, According to Pregnancy Outcome (Wanted Percentages)

| Pregnancy Outcome | Wanted/Not Wanted Consistent | Wanted/Not Wanted Inconsistent | Total |
|--|------------------------------|--------------------------------|------------|
| Live birth | 25.9 | 22.5 | 153 |
| Abortion | 28.3 | 21.3 | 140 |
| Stillbirth, miscarriage, ectopic pregnancy | 32.5 | 27.5 | 240 |
| Total | 26.9 | 23.8 | 333 |

TABLE 1b
Consistency of Women's First Wanted/Not Wanted Status Between Time 1 and Time 2 About 18 Months Postpartum, According to Pregnancy Outcome (Wanted Percentages)

| Pregnancy Outcome | Wanted/Not Wanted Consistent | Wanted/Not Wanted Inconsistent | Total |
|--|------------------------------|--------------------------------|------------|
| Live birth | 25.0 | 25.0 | 150 |
| Abortion | 28.0 | 22.0 | 130 |
| Stillbirth, miscarriage, ectopic pregnancy | 33.0 | 25.0 | 210 |
| Total | 28.0 | 24.0 | 300 |

rately for different Time 1 intention statuses. As a result, in those two tables the dependent variable can capture shifts toward more positive and more negative views. The dependent variable comprises three categories: More positive, more negative, and consistent. A more positive report resulted (1) if the woman claimed

at Time 1 that the pregnancy was unwanted, and then reported at Time 2 that the pregnancy was mistimed or intended, or (2) if she said at Time 1 that the pregnancy was mistimed and at Time 2 that it was intended. A more negative report resulted (1) if the respondent said at Time 1 that the pregnancy was intended and said

ion ▾ Processing and Analysis ▾ Command Files Search

| Name | Date modified |
|-----------------------|-------------------|
| 00_masterdo | 4/27/2016 3:34 PM |
| 01_run_extract_hhcomp | 4/27/2016 2:30 PM |
| 02_recode_hhcomp | 4/27/2016 2:30 PM |
| 03_run_extract | 4/27/2016 2:30 PM |
| 04_recode | |
| 05_run_hier_extract | |
| 06_recode_who | |
| 07_recode_activity | |
| 08_recode | |
| 09_asr_r8r | |

- 00_masterdo
- 01_run_extract_hhcomp
- 02_recode_hhcomp
- 03_run_extract
- 04_recode
- 05_run_hier_extract
- 06_recode_who
- 07_recode_activity
- 08_recode
- 09_asr_r8r

```

26 /* Survey set the data to incorporate PSU and strata*/
27 svyset [pweight=peravg], strata(sect) pau(sect)
28
29
30 /* Table 1 Descriptives—Overall sample*/
31 /* 1995 vs. 2006-10 women descriptives (col's 1 and 2 in T1, descriptives)*/
32 by: survey1: duration ucon2 ucon3 cohbirth black hispan other dadedu3 dadedu2 dadedu1 dadedu4 ///
33 wanted1 wanted2 wanted3 wanted4 instace church ///
34 lths high somecoll coll agency apgedady planned momeprekida priormar priorcoh anyprioruni ///
35 mistimed unwanted ageoldest1 ageoldest2 ageoldest3 babbyy priorunikids timebaby anyshotgun ///
36 if bl=1 [weight=peravg]
37
38 /*Table 2: Logistic regressions*/
39 /*Model 1*/
40 local altvmar "altvmar"
41 foreach j in 1995 2007 {
42 svy: logit tvsplit tvduration cohunion cohbirth "altvmar" ///
43 if bl=1 & survey=j, or
44
45 /*Test coefficients in Table 3*/
46 /* Row 1*/

```

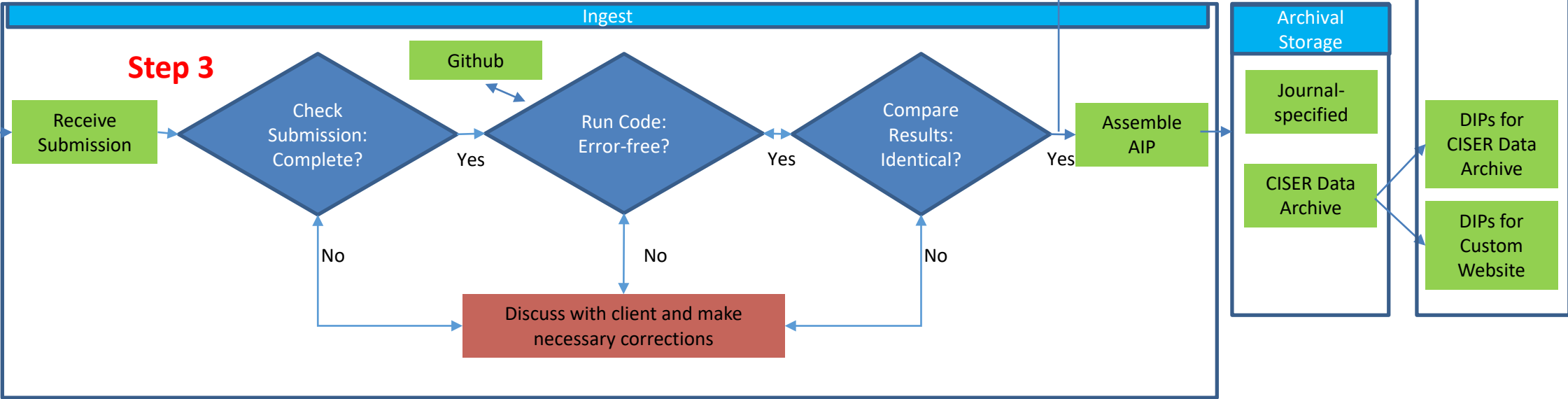
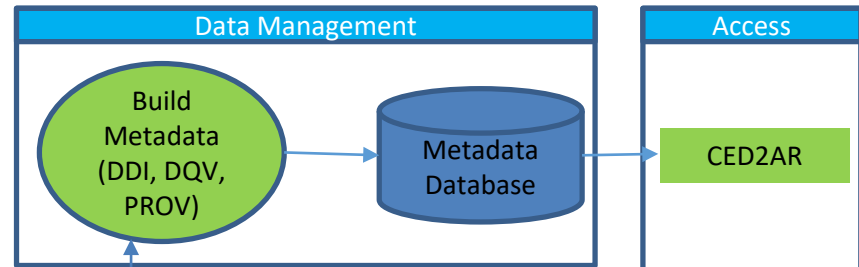
Variables Manager

| # | Name | Label | Type | Format | Value label | Notes |
|---|------------|-----------------------------------|--------|--------|---------------|-------|
| | rectype | Record Type | byte | %8.0g | | |
| | caseid | ATUS Case ID | double | %14.0f | | |
| | year | Survey year | long | %12.0g | | |
| | pernum | Person number (general) | byte | %8.0g | pernum_bl | |
| | lineno | Person line number | int | %21.0g | lineno_bl | |
| | month | Month of ATUS interview | int | %21.0g | month_bl | |
| | day | ATUS interview day of the week | byte | %21.0g | day_bl | |
| | wt06 | Person weight, 2006 methods... | double | %17.0f | | |
| | age | Age | int | %6.0g | | |
| | sex | Sex | byte | %21.0g | sex_bl | |
| | race | Race | int | %42.0g | race_bl | |
| | marst | Marital status | byte | %24.0g | marst_bl | |
| | educ | Highest level of school comple... | int | %47.0g | educ_bl | |
| | empstat | Labor force status | byte | %22.0g | empstat_bl | |
| | spousepres | Spouse or unmarried partner i... | byte | %30.0g | spousepres_bl | |
| | wb_resp | Well-Being Module Respondent | byte | %22.0g | wb_resp_bl | |
| | wbvnt | Well-being Module final statu... | float | %15.6f | wbvnt_bl | |

```

p1it]cohunion+(t2007_tvsplit)'altv
p1it] 'altvmar'=(t1995_tvsplit]cohah
p1it]cohunion+(t2007_tvsplit]cohah
p1it]cohbirth
p1it]cohbirth
t2007_tvsplit]cohbirth +(t2007_tv

```





COMPUTE TOTALINCOME = INCOME1 + INCOME2 + INCOME3.

COMPUTE TOTALINCOME = SUM(INCOME1, INCOME2, INCOME3).

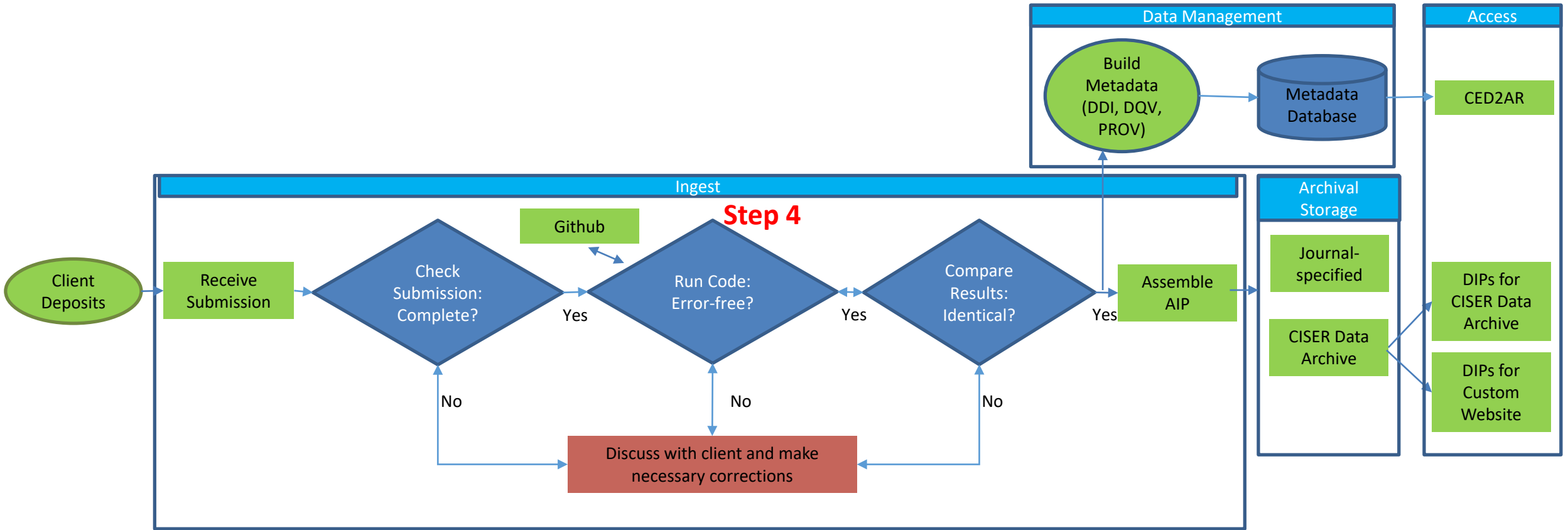




Table 1. Socio-demographic characteristics of parents (N individuals = 12,163)

| | Full Sample | Mothers | Fathers | |
|--|-----------------|-----------------|-----------------|----------------|
| N (individuals) | 12163 | 7232 | 4931 | actP |
| Age | 38.38 (9.21) | 36.91 (9.26) | 40.22 (8.76) | age z2agesd |
| Number of children in the HH (reference = 1 child) | 40.05 | 41.97 | 37.67 | nkidhh31 |
| 2 children | 39.50 | 39.18 | 39.91 | nkidhh32 |
| 3+ children | 20.44 | 18.85 | 22.43 | nkidhh33 |
| Age of youngest child (reference = <6) | 47.53 | 47.68 | 47.33 | ychild1 |
| 6-12 | 32.83 | 31.71 | 34.22 | ychild2 |
| 13-17 | 19.65 | 20.61 | 18.44 | ychild3 |
| Race/ethnicity (reference = White, non-Hispanic) | 63.44 | 62.05 | 65.16 | hr1 |
| Black, non-Hispanic | 9.88 | 11.72 | 7.60 | hr2 |
| Hispanic | 20.11 | 20.43 | 21.17 | hr3 |
| Other, non-Hispanic | 6.57 | 5.80 | 7.53 | hr4 |
| Spouse or partner in the HH | 82.80 | 74.10 | 93.64 | pinhh |
| Household income (reference = <\$25,000) | 18.00 | 21.32 | 13.87 | inccat1 |
| \$25,000-\$99,999 | 57.70 | 56.32 | 59.42 | inccat2 |
| \$100,000 + | 23.14 | 20.78 | 26.08 | inccat3 |
| Income missing | 1.16 | 1.58 | 0.64 | inccat4 |
| Employment status (reference = full time) | 62.43 | 44.69 | 84.51 | ws1 |
| Part-time work | 14.32 | 21.28 | 5.67 | ws2 |
| No paid work | 23.24 | 34.03 | 9.82 | ws3 |
| College degree + | 35.01 | 34.86 | 35.19 | collgrad |
| Enrolled in school | 5.82 | 7.77 | 3.39 | mschool |

Notes: 2010, 2012, and 2013 ATUS well-being sample, men and women with children in the household. N's are unweighted; means/percentages are weighted (using individual weights for individual-level and activity weights for activity-level). Standard deviations in parentheses. HH = household.
 * = Mothers' characteristics are significantly different (p<.05) from fathers'. (†est`v`)

```

**print Table 1
disp `varname|fs|men|women||diff m v. w'
varname|fs|men|women||diff m v. w
disp `actP|nt`||`nt1`||`nt2`"
actP|12163|4931|7232

foreach v in `alvars' {
  disp "`v'|`v'f'|`v'1'|`v'2'|`test`v'"
}

age|38.4752||40.3477||36.9548||0
collgrad|.3478||.3479||.3478||.9938
hr1|.616400000000000001||.6424000000000001||.5954||.0001
hr2|.1033||.0772||.1245||0
hr3|.2131||.2117||.2142||.8134
hr4|.0672||.0688||.0659||.6115
inccat1|.1848||.1364||.224||0
inccat2|.5628000000000001||.5862000000000001||.5439000000000001||.0002
inccat3|.2421||.2691||.2202||0
inccat4|.0102||.0082||.0119||.1349
inschool|.0568||.0338||.0756||0
nkidhh31|.4048||.3895||.4173||.0152
nkidhh32|.3886||.3929||.3851||.4867
nkidhh33|.2066||.2176||.1976||.0379
pinhh|.8261000000000001||.9323||.7399||0
ws1|.6147||.8238000000000001||.445||0
ws2|.1418||.0624||.2062||0
ws3|.2435||.1138||.3488||0
ychild1|.4681||.464||.4713||.5266000000000001
ychild2|.3287||.3334||.3248||.425
ychild3|.2033||.2026||.2038||.9012

foreach v in age {
  disp "z`v`sd|(`v`sd')||(`v`sd1')||(`v`sd2'"
}

z2agesd|(9.08)||(.8630000000000001)||(.9140000000000001)
} //end T1
  
```

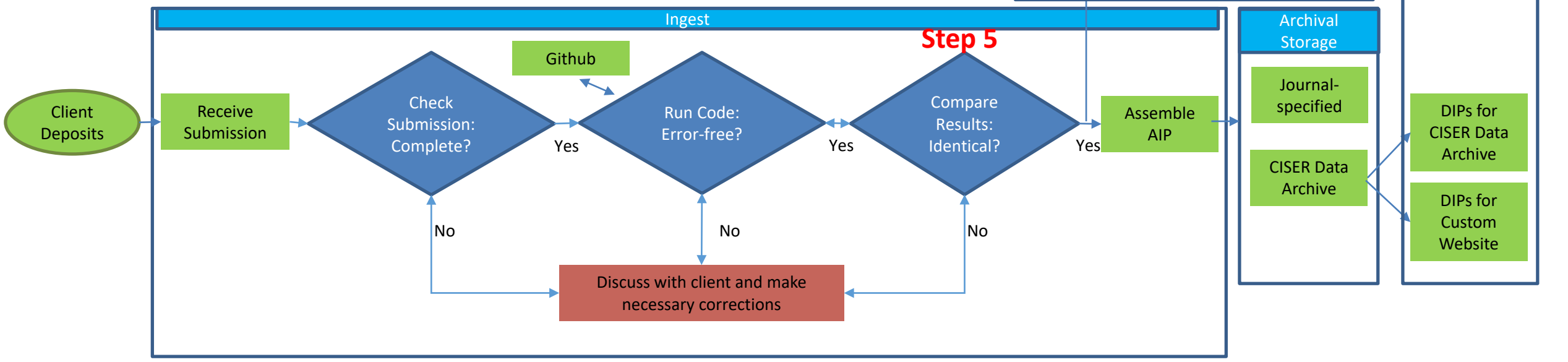




Table 1. Sociodemographic characteristics of parents (N = 12,163)

| | Full Sample | Mothers | Fathers | actP |
|--|--------------|--------------|--------------|----------|
| N (individuals) | 12163 | 7232 | 4931 | |
| Age | 38.38 (9.21) | 36.91 (9.26) | 40.22 (8.76) | age |
| Number of children in the HH (reference = 1 child) | 40.05 (9.21) | 41.97 (9.26) | 37.67 (8.76) | nkidhh31 |
| 2 children | 39.50 | 39.18 | 39.91 | nkidhh32 |
| 3+ children | 20.44 | 18.85 | 22.43 | nkidhh33 |
| Age of youngest child (reference = <6) | 47.53 (9.21) | 47.68 (9.26) | 47.33 (8.76) | ychild1 |
| 6-12 | 32.83 | 31.71 | 34.22 | ychild2 |
| 13-17 | 19.65 | 20.61 | 18.44 | ychild3 |
| Race/ethnicity (reference = White, non-Hispanic) | 63.44 (9.21) | 62.05 (9.26) | 65.16 (8.76) | hr1 |
| Black, non-Hispanic | 9.88 | 11.72 | 7.60 | hr2 |
| Hispanic | 20.11 | 20.43 | 19.71 | hr3 |
| Other, non-Hispanic | 6.57 | 5.80 | 7.53 | hr4 |
| Spouse or partner in the HH | 82.80 | 82.61 | 83.99 | pinhh |
| Household income (reference = <\$25,000) | 18.00 (9.21) | 21.32 (9.26) | 13.87 (8.76) | inccat1 |
| \$25,000-\$99,999 | 57.70 | 56.28 | 59.42 | inccat2 |
| \$100,000 + | 23.14 | 24.21 | 26.08 | inccat3 |
| Income missing | 1.16 | 1.02 | 0.64 | inccat4 |
| Employment status (reference = full time) | 62.43 (9.21) | 44.69 (9.26) | 84.51 (8.76) | ws1 |
| Part-time work | 14.32 | 14.18 | 15.67 | ws2 |
| No paid work | 23.24 | 24.35 | 9.82 | ws3 |
| College degree + | 35.01 (9.21) | 34.86 (9.26) | 35.19 (8.76) | collgrad |
| Enrolled in school | 5.82 | 5.68 | 3.39 | mschool |

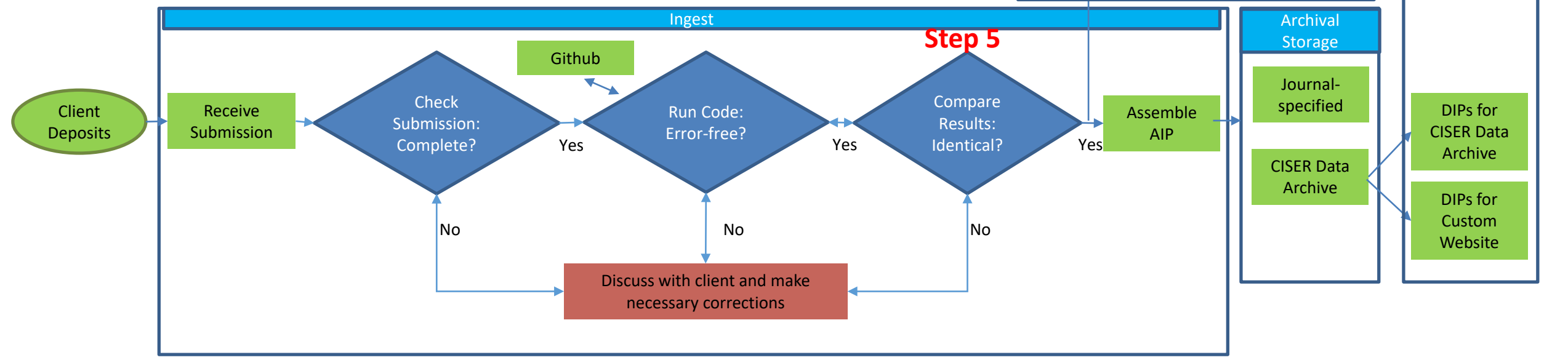
Notes: 2010, 2012, and 2013 ATUS well-being sample, men and women with children in the household. N's are unweighted; means/percentages are weighted (using individual weights for individual-level and activity weights for activity-level). Standard deviations in parentheses. HH = household.
* = Mothers' characteristics are significantly different (p<.05) from fathers'.

```

**print Table 1
disp _varname|fs|men|women|diff m v. w
_varname|fs|men|women|diff m v. w
disp _actP|nt|nt1|nt2
_actP|12163|4931|7232

foreach v in age {
disp "zz'v|sd|('v|sd')|('v|sd1')|('v|sd2')
} //end T1

```

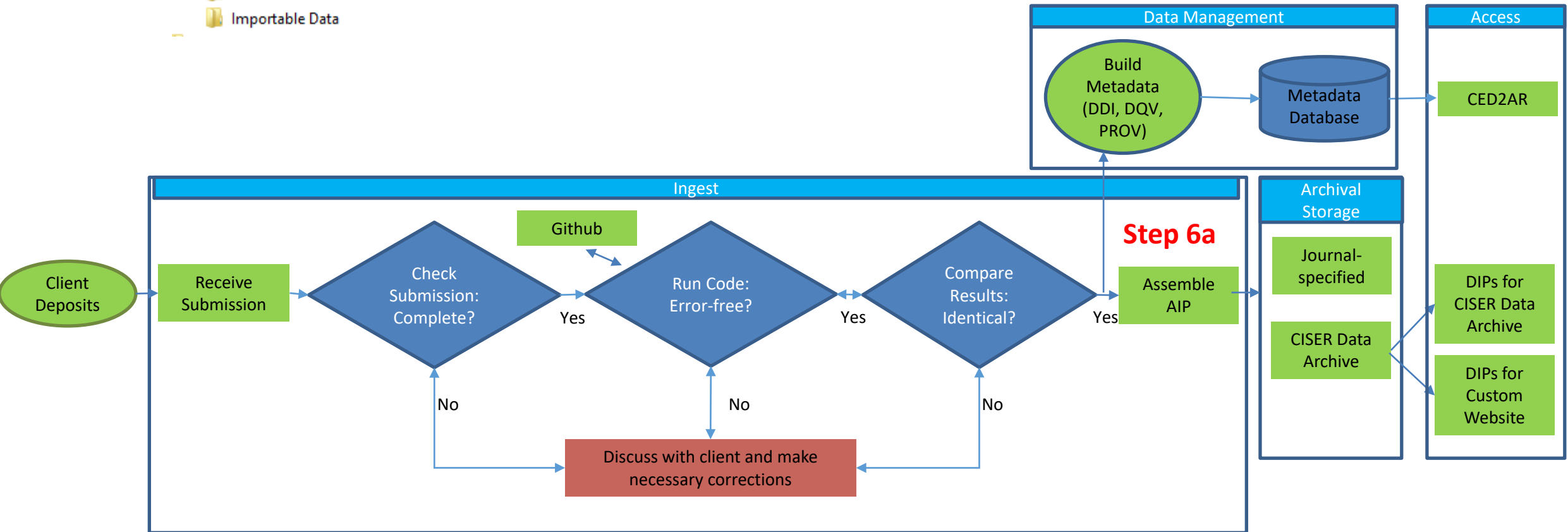


Teaching Integrity in Empirical Research (TIER) protocol

<https://www.haverford.edu/project-tier/protocol-v2>



- Replication Documentation
 - Original Data and Metadata
 - Metadata
 - Original Data
 - Processing and Analysis
 - Analysis Data
 - Command Files
 - Importable Data

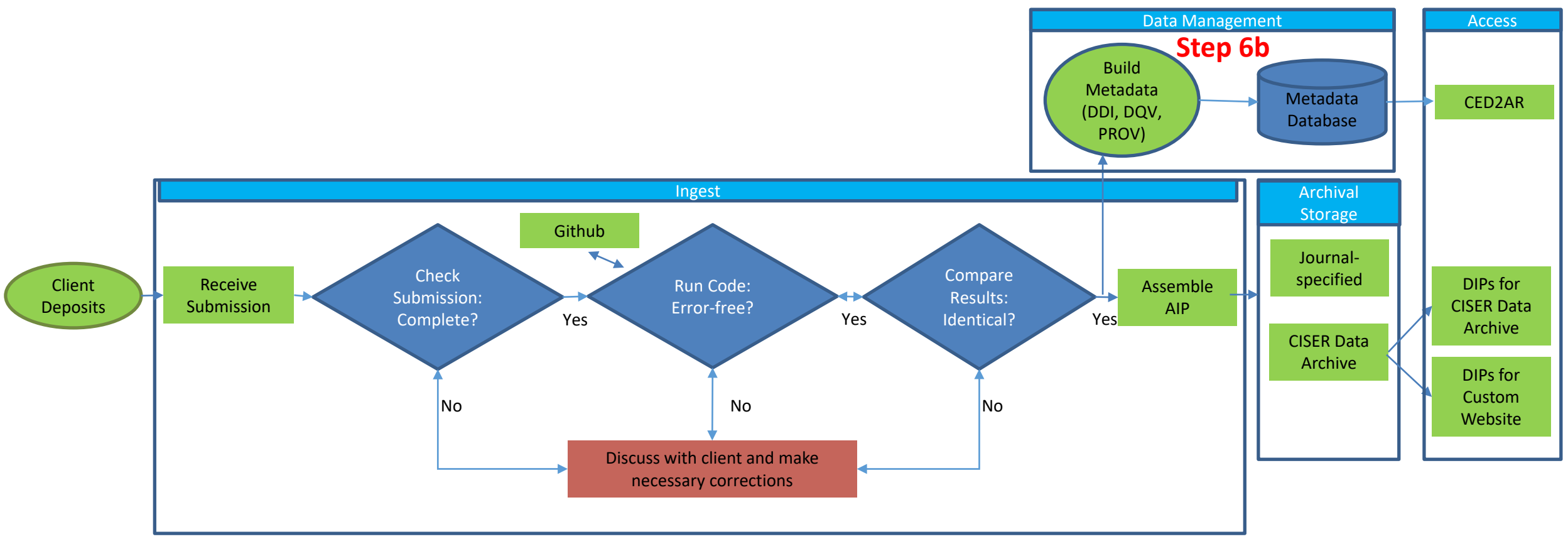


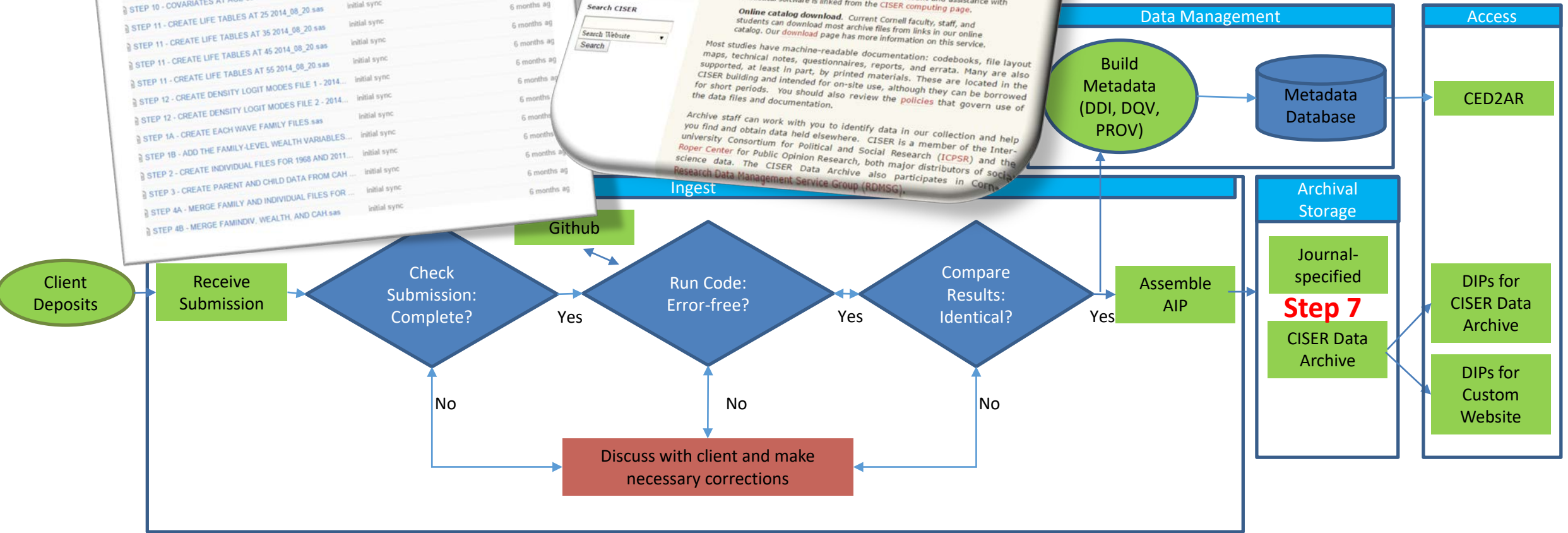
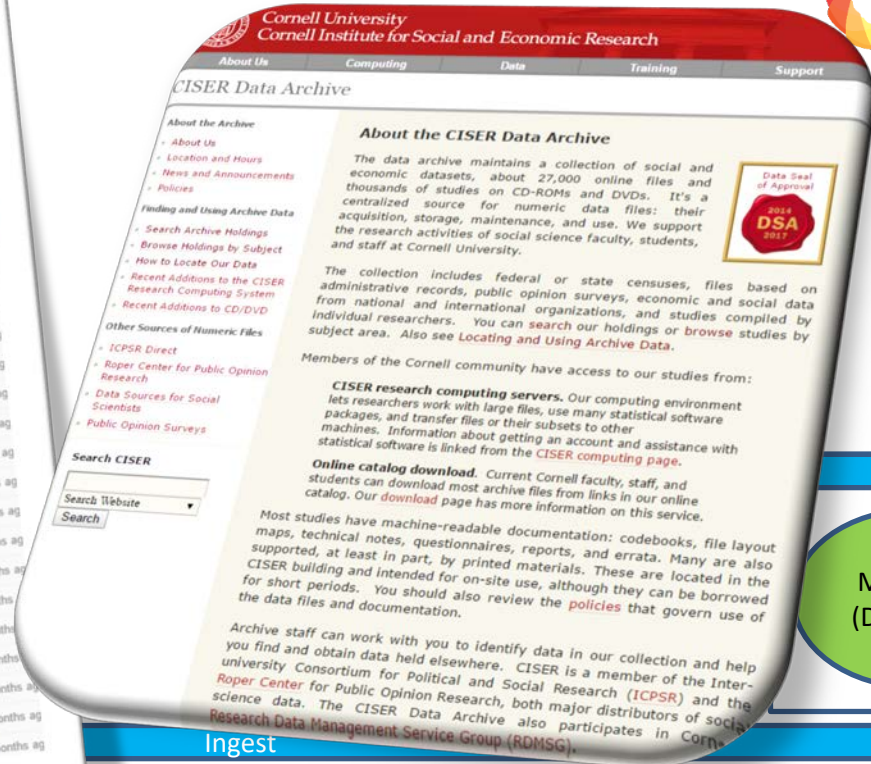
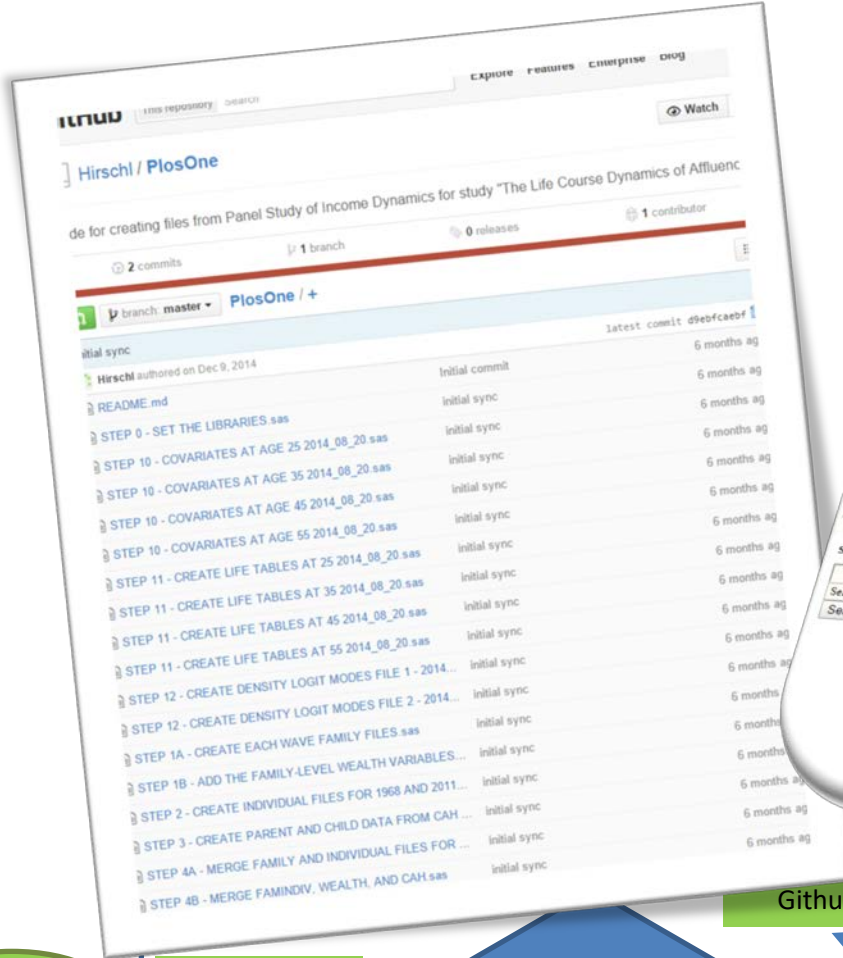


Comprehensive Extensible Data Documentation and Access Repository (CED²AR)

Data Quality Vocabulary (DQV): <http://w3c.github.io/dwbp/vocab-dqg.html>

PROV: <http://www.w3.org/TR/prov-primer/>







CISER Data Archive: Online Catalog

About the Archive

- About Us
- Location and Hours
- News and Announcements
- Policies

Finding and Using Archive Data

- Search Archive Holdings
- Browse Holdings by Subject
- How to Locate Our Data
- Recent Additions to the CISER Research Computing System
- Recent Additions to CD/DVD

Other Sources of Numeric Files

- ICPSR Direct
- Roper Center for Public Opinion Research
- Data Sources for Social Scientists
- Public Opinion Surveys

Search CISER

Search Website:
Search

How parents fare: Mothers' and fathers' subjective well-being in time with children

Bibliographic Information:

Musick, Kelly, Ann Meier, and Sarah Flood (Forthcoming, 2016). American Sociological Review [producer]. Codebook: R2-MUSICK-2016. This study includes files created by Cornell researchers and/or staff.

View Abstract

File Information:

Program files and data

| Type of File | Directory \ File Name | Size / Size Zipped |
|----------------|--|--------------------|
| Zipped Archive | V:\r2\MUSICK-2016\ASR-041416-REV061616.zip | 688 KB / 77 KB |

Notes: Uses Stata, see _readme.txt for instructions.

Output tables

| Type of File | Directory \ File Name | Size / Size Zipped |
|-------------------|---|--------------------|
| Excel Spreadsheet | V:\r2\MUSICK-2016\mmf_2016.06.16_ASR.xlsx | 902 KB / 723 KB |

Abstract: The shift to more time-intensive and child-centered parenting in the U.S. is widely assumed to be positively linked to healthy child development, but implications for adult well-being are less clear. We assess multiple dimensions of parents' subjective well-being in activities with children and explore how the gendered nature of time potentially contributes to differences in mothers' and fathers' parenting experiences. Relying on nationally representative time diary data linked to respondents' feelings in activities from the 2010, 2012, and 2013 well-being module of the American Time Use Survey (N = 12,163 persons and 38,036 activities), we find that parents consistently report greater subjective well-being in activities with children than without. Mothers, however, report less happiness, more stress, and greater fatigue in time with children than fathers. These gaps are relatively small and can be accounted for by differences in the activities that mothers and fathers engage in with children, whether other adults are present, and the quality of their sleep and leisure. We go beyond prior work on parental happiness and life satisfaction to document how contemporary parenting is woven differently into the lives of mothers and fathers.

Keywords: parenting, subjective well-being, gendered family roles, time use

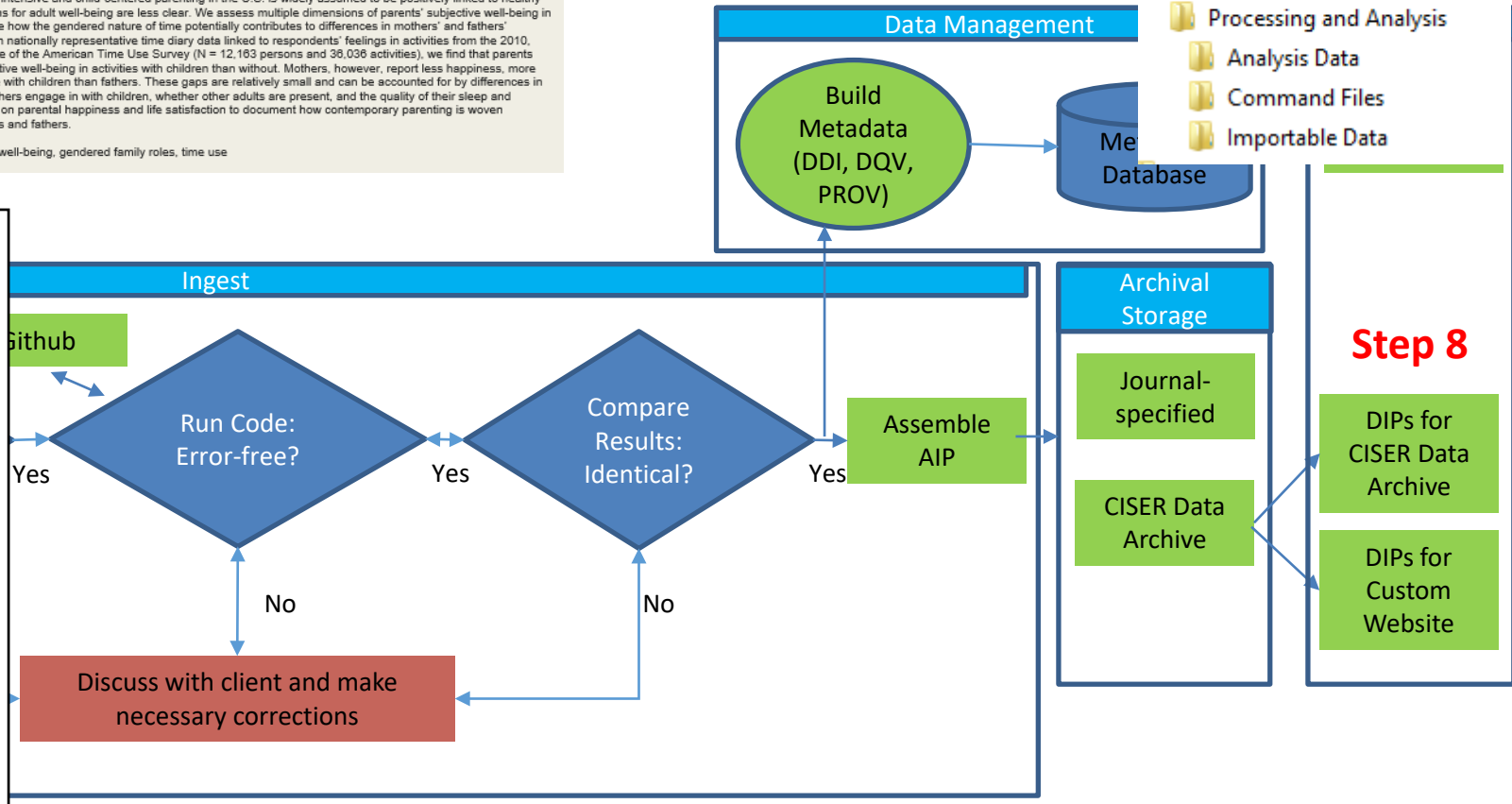
Links to:

- Metadata with variable level search through CISER's instance of CED2AR
- Journal article
- Replication materials:
V:\R2\<<1stAuth>\<r2no>\



- Replication Documentation
- Original Data and Metadata
- Metadata
- Original Data
- Processing and Analysis
- Analysis Data
- Command Files
- Importable Data

- Will also include:
- Persistent URL:
<http://doi.org/10.6077/J5CISER2775>
 - Study citations
 - Data citation
 - Documentation about known quality issues
 - User Feedback Form
 - Forum or mailing list
 - Social Media Accounts (e.g., Facebook, Twitter)

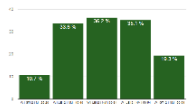




About the Book

The United States has been epitomized as a land of opportunity, where hard work and skill can bring about personal success and economic well-being. Yet in their pursuit of the American Dream, many will experience poverty firsthand. Just how real that risk can be found in our new poverty risk calculator.

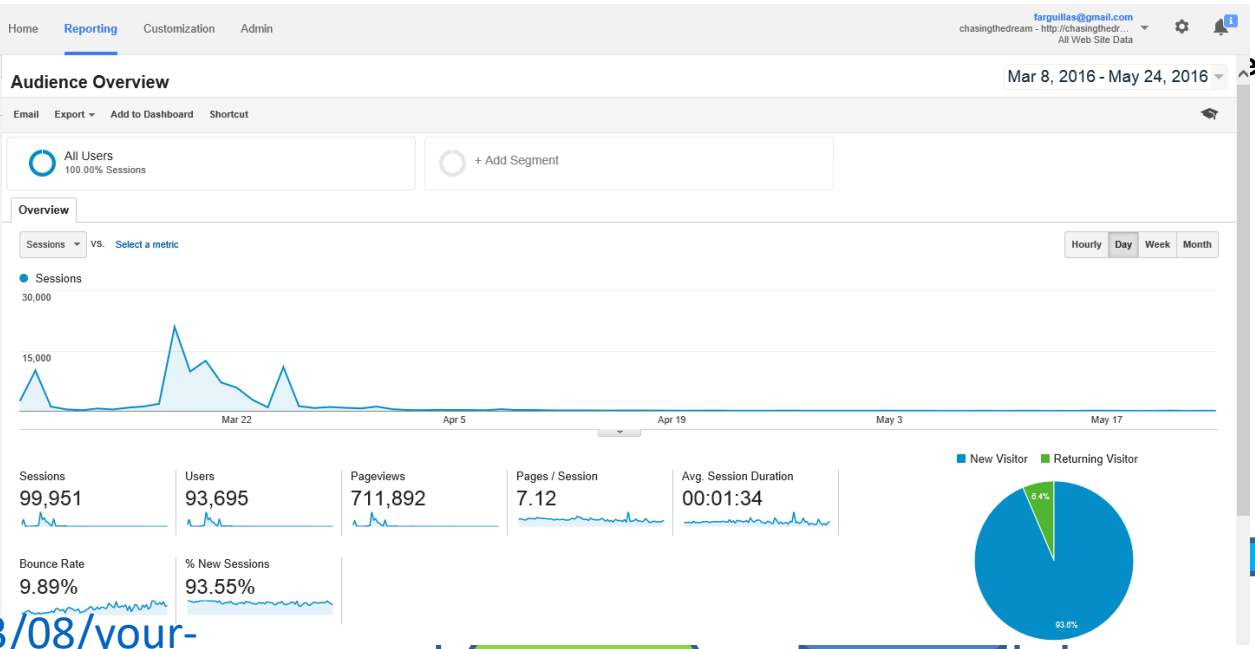
Risk Calculator



Have you ever wondered what your risk of poverty might be in the future? Our calculator will estimate your personal risk over the next 5, 10, or 15 years.

About the Authors

- Mark Robert Rank, Washington University
- Thomas A. Hirschl, Cornell University
- Kirk A. Foster, University of South Carolina

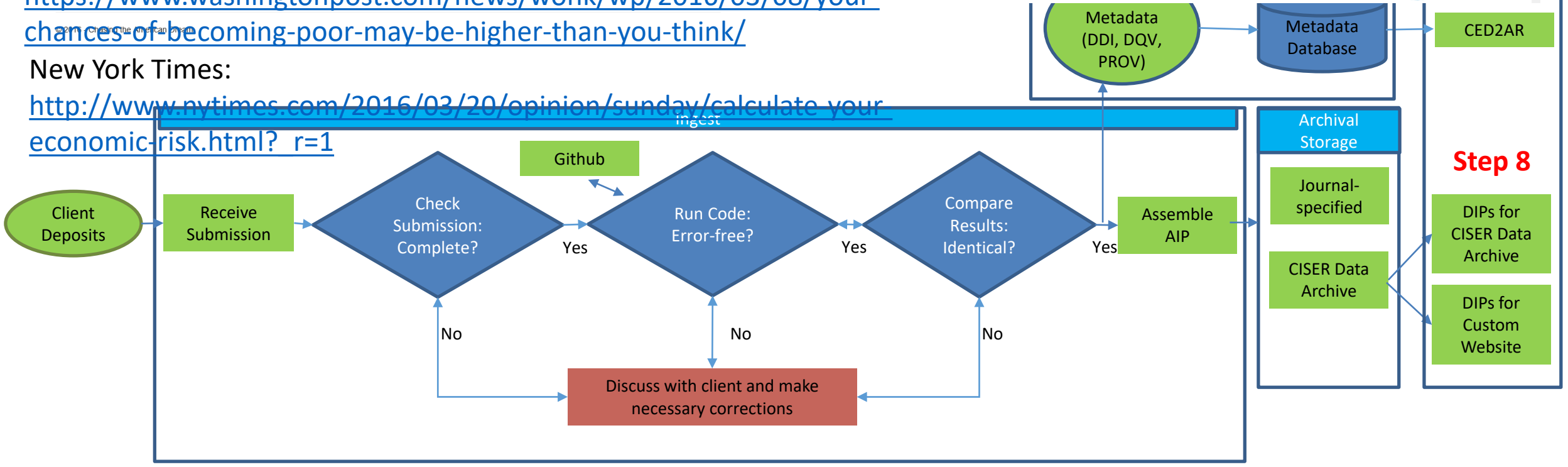


Washington Post:

<https://www.washingtonpost.com/news/wonk/wp/2016/03/08/your-chances-of-becoming-poor-may-be-higher-than-you-think/>

New York Times:

<http://www.nytimes.com/2016/03/20/opinion/sunday/calculate-your-economic-risk.html? r=1>





Conclusion

- Process is time-intensive, costly, and confusing if data and code quality is low
- Suggest code review, use of versioning software, adherence to TIER protocol in managing files
- Reach out to Cornell graduate students and provide training on how to prepare research materials for reuse thru classes or workshops.
- CURATE AS YOU CODE AND CODE WITH REUSE IN MIND.
- For archives:
 - Incorporate DQV and PROV
 - Follow TOP Guidelines, Open Data Certificate Requirements, and TIER Protocol for managing and packaging replication materials
 - Anything else?



Thank you!

- Contact author: Florio Arguillas (foa2@cornell.edu)

