CHRISTOPHER PRENER, PH.D.
SPRING, 2017

WEEK 07
LECTURE 07

INTRO TO GIS$_C$

GEODATABASES

CHRISTOPHER PRENER, PH.D.
SPRING, 2017

WEEK 07
LECTURE 07

INTRO TO GIS$_C$

POTPOURRI

AGENDA

1. Follow-up
2. Advanced GitHub
3. Creating New Variables
4. Geodatabases

1 FOLLOW-UP

GISc WORKFLOW

tasks for each step:

1. planning
2. data collection
2. organization
3. documentation
4. execution

data storage

CSV

raw data

STATA

data cleaning

geodatabase

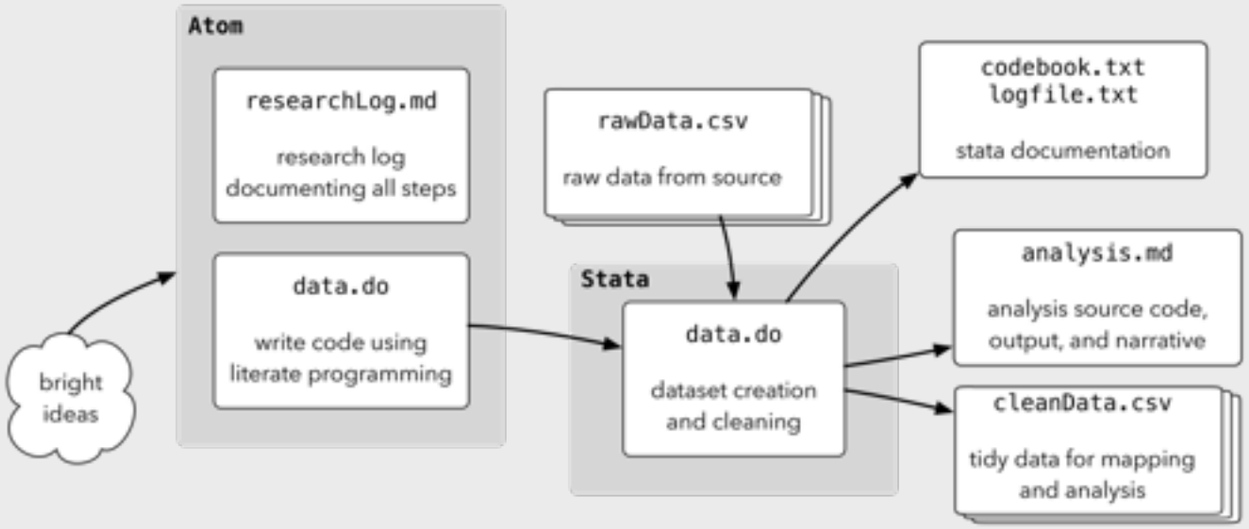
ArcGIS

data visualization

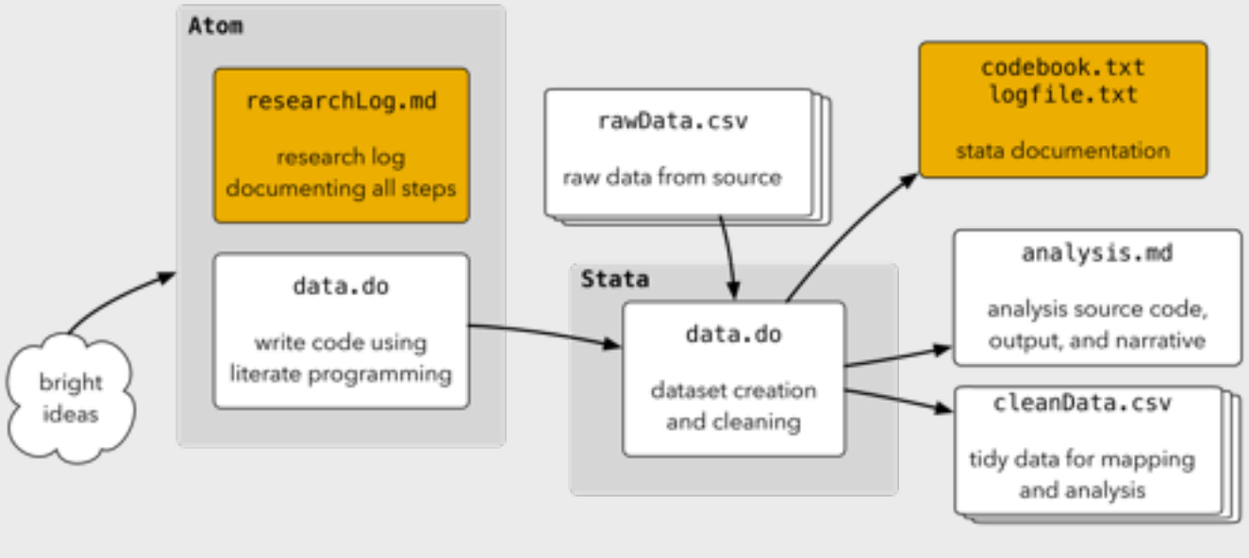
PDF

dissemination

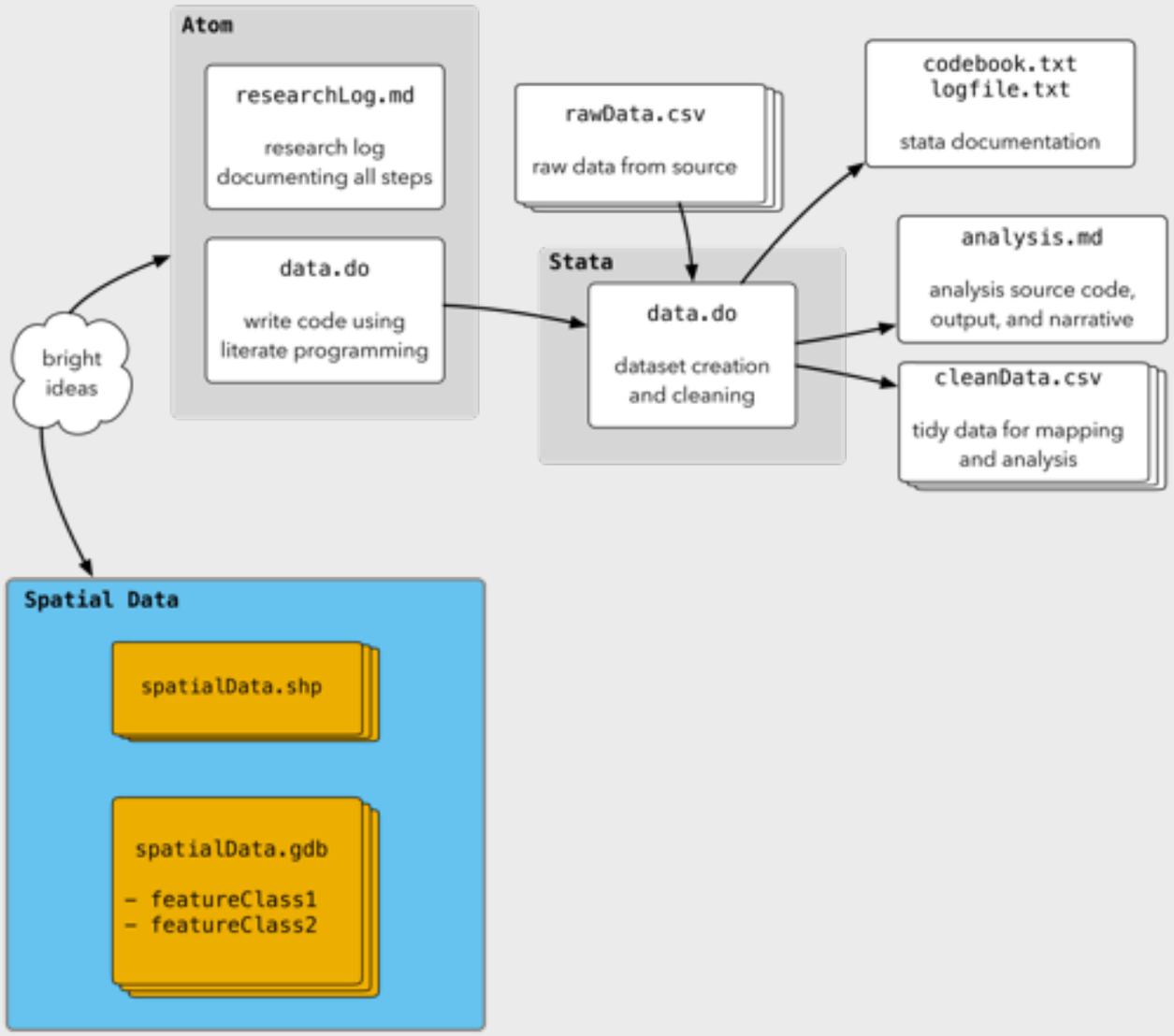
GISc WORKFLOW



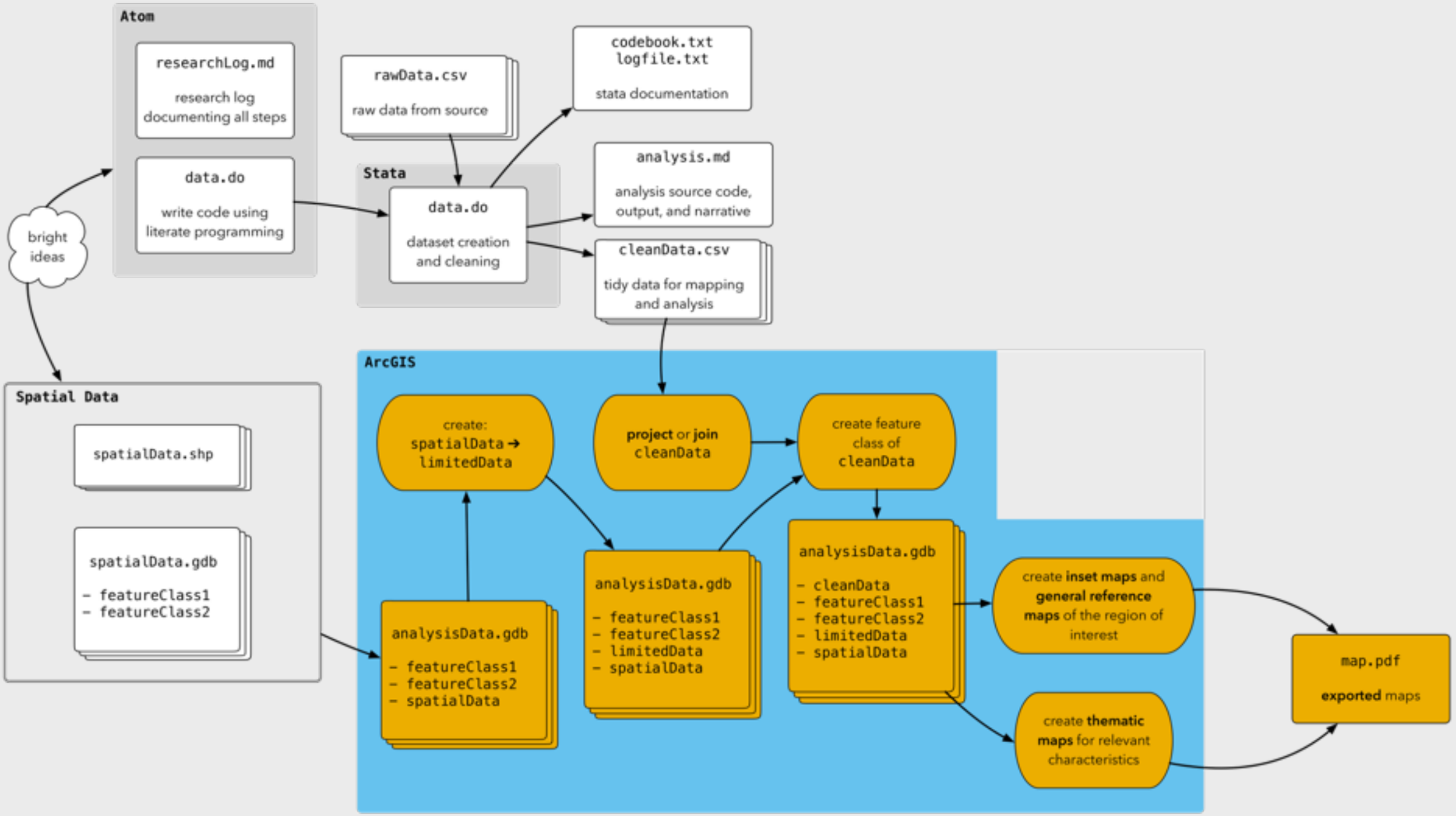
GISc WORKFLOW



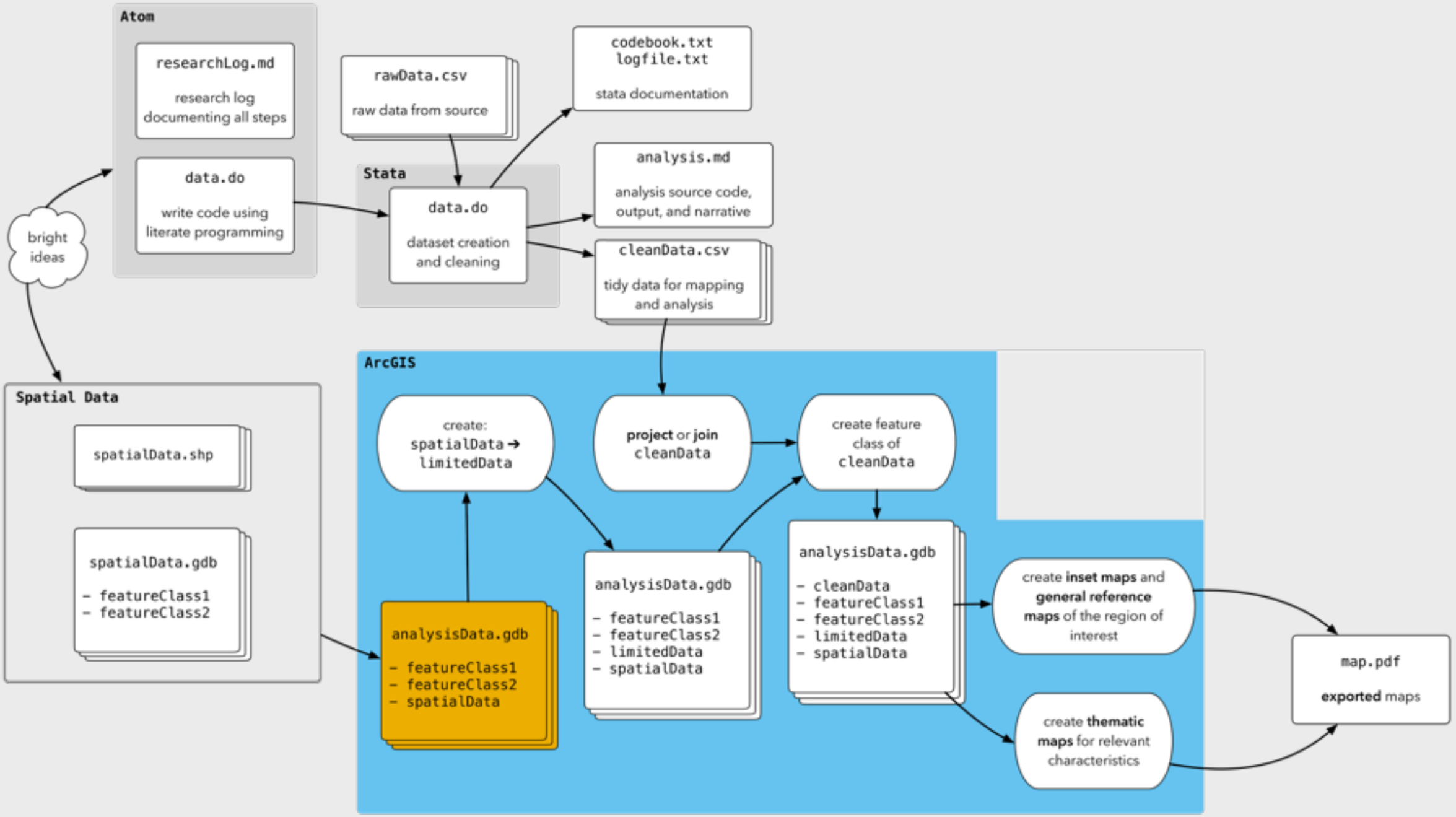
GISc WORKFLOW



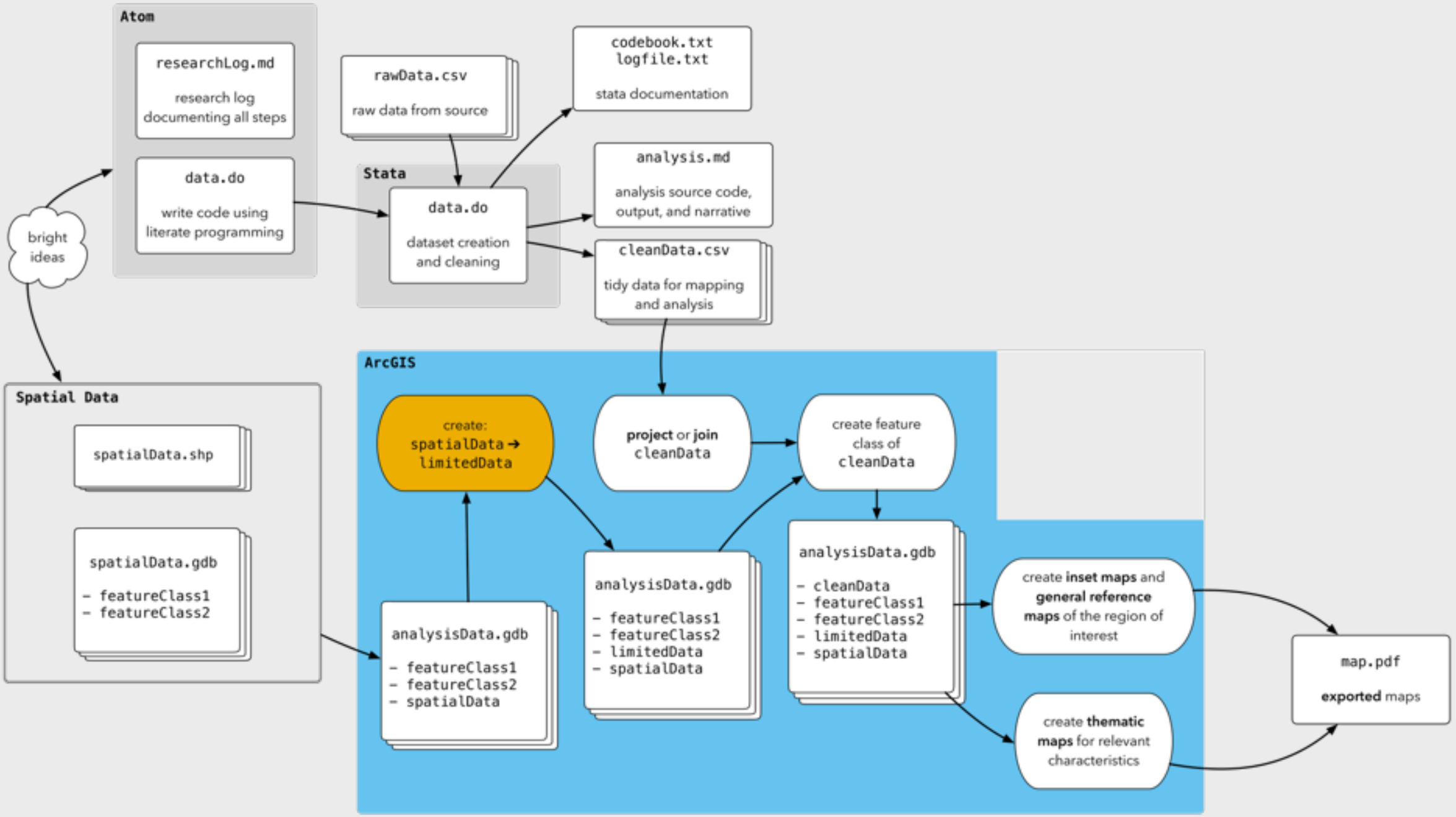
GISc WORKFLOW



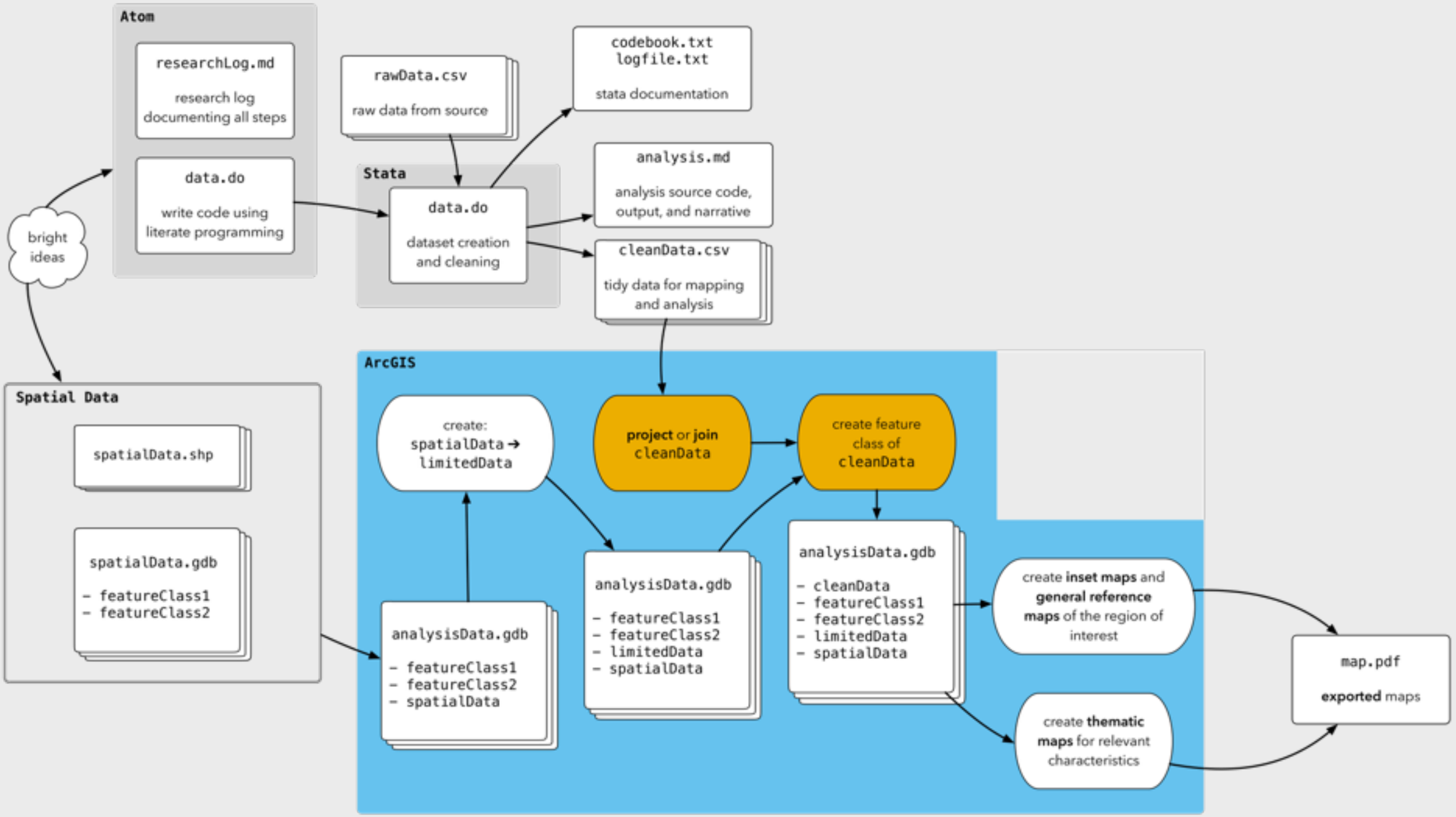
GISc WORKFLOW



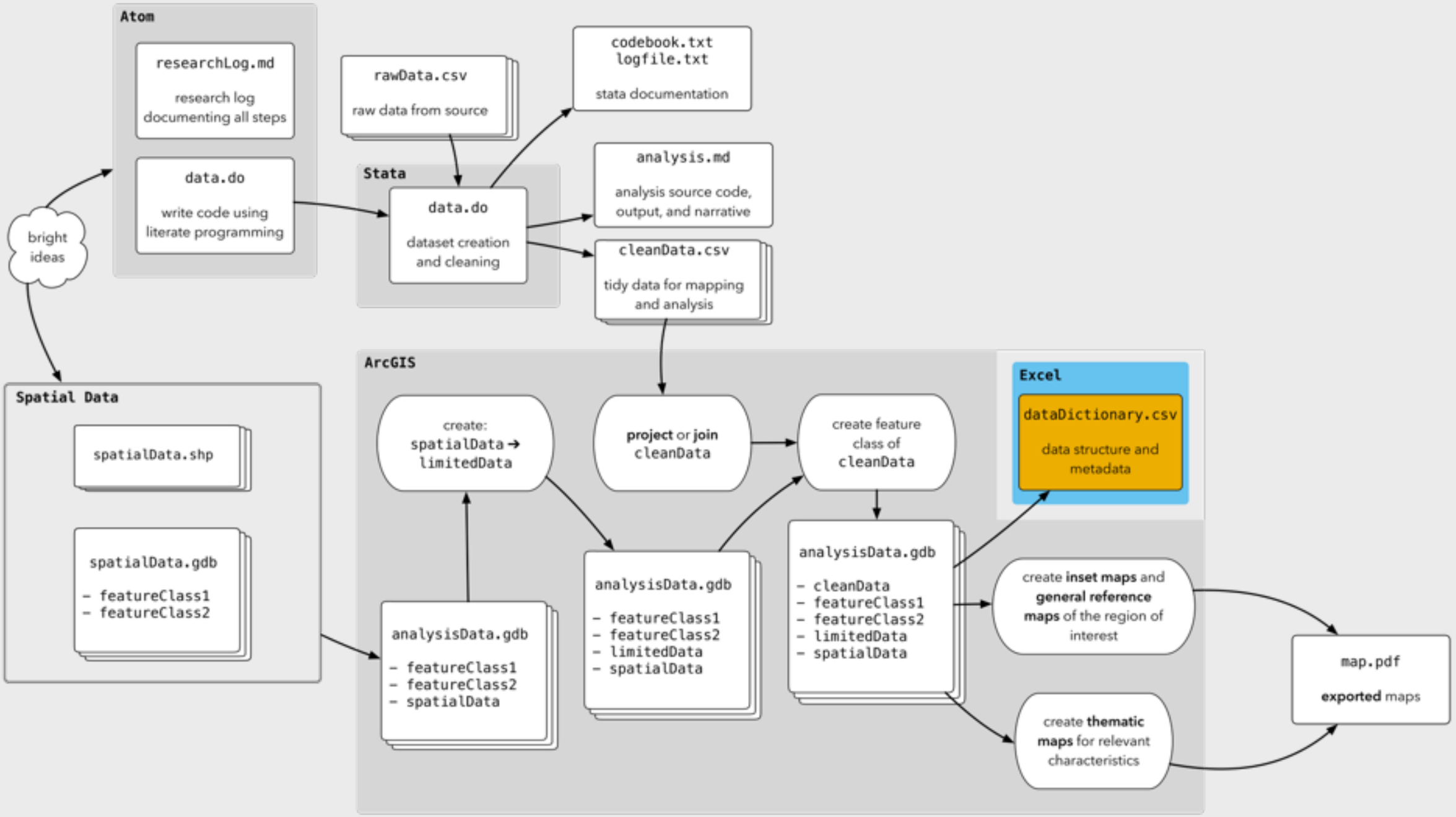
GISc WORKFLOW



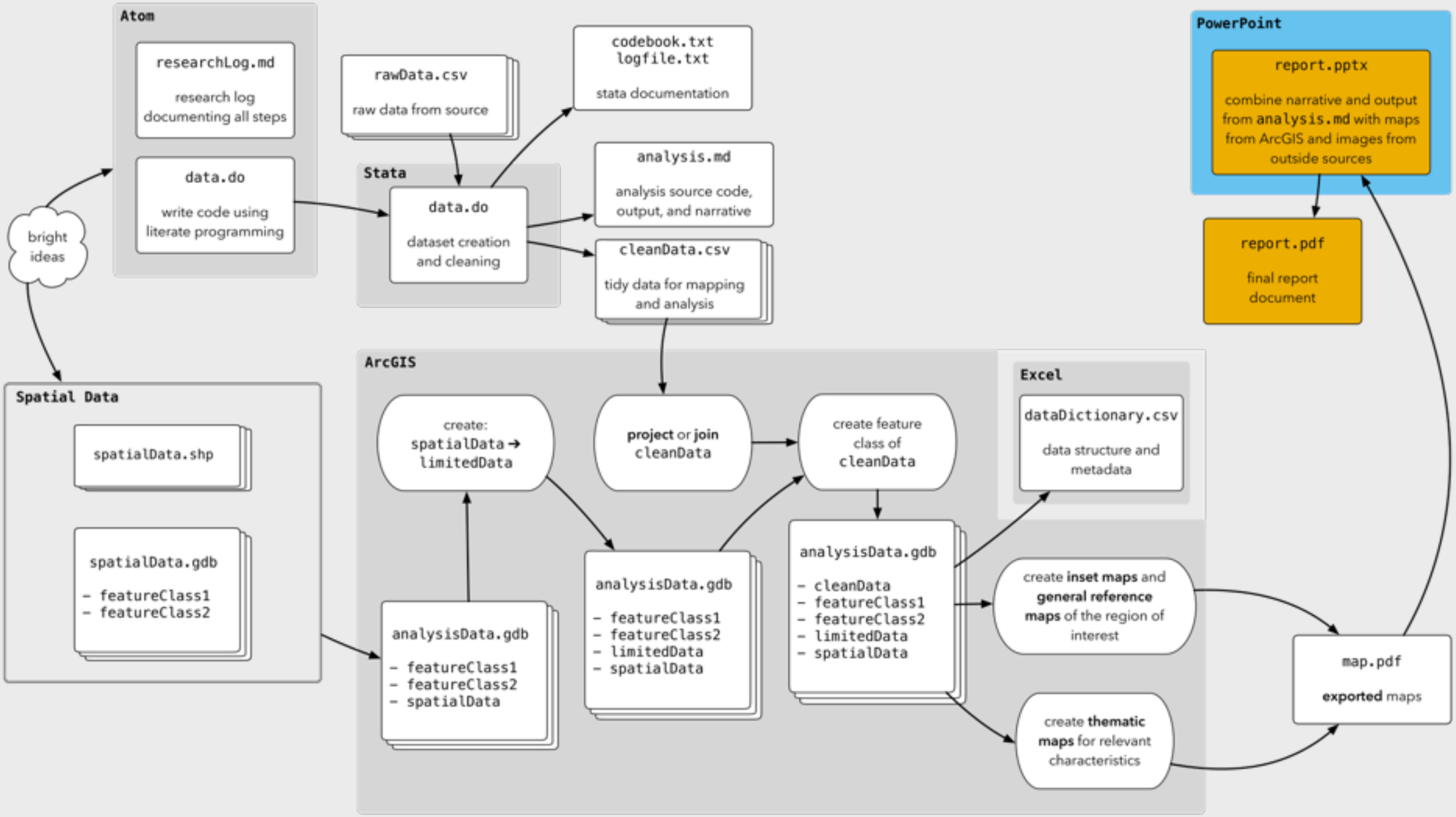
GISc WORKFLOW



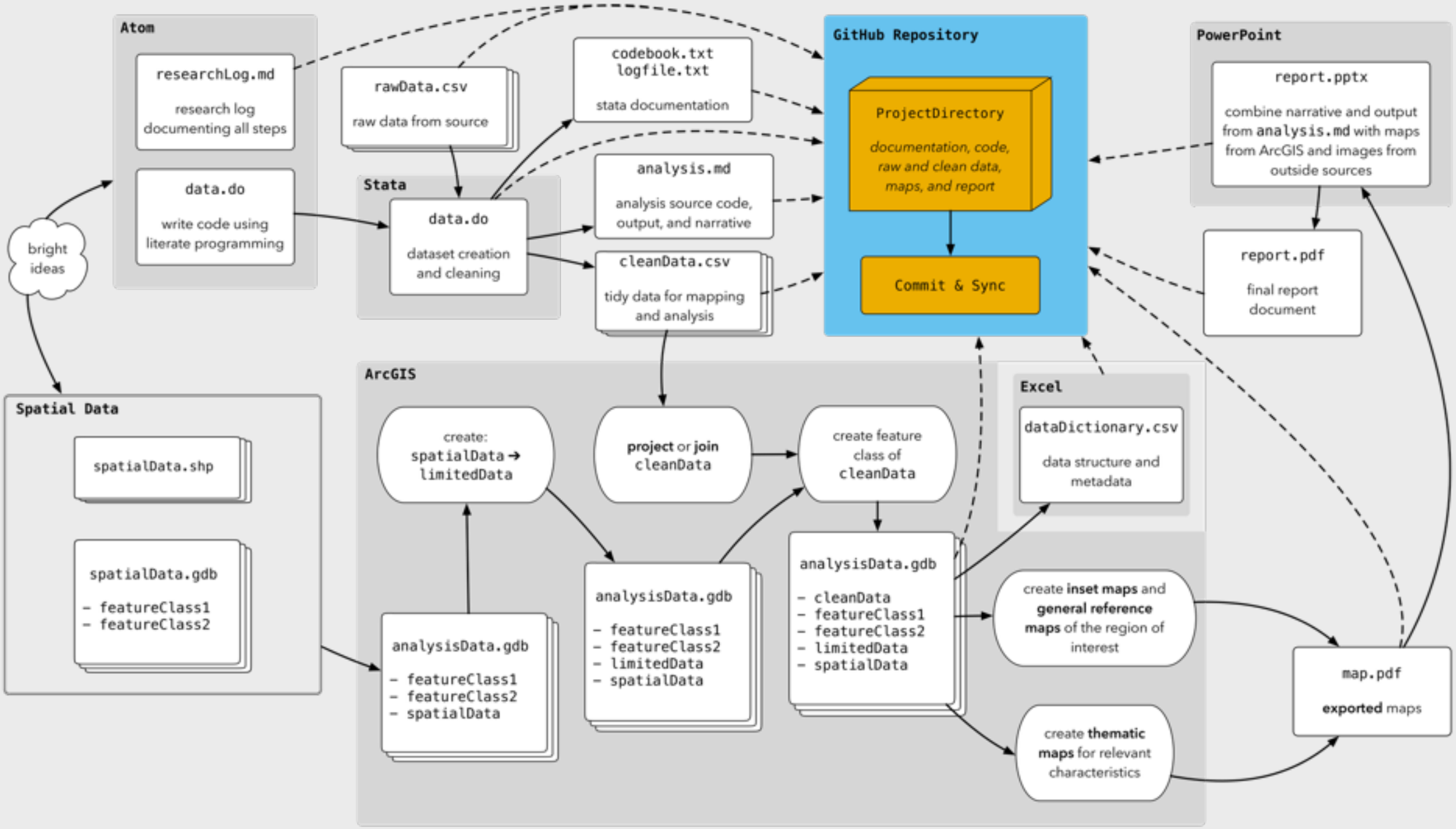
GISc WORKFLOW



GISc WORKFLOW

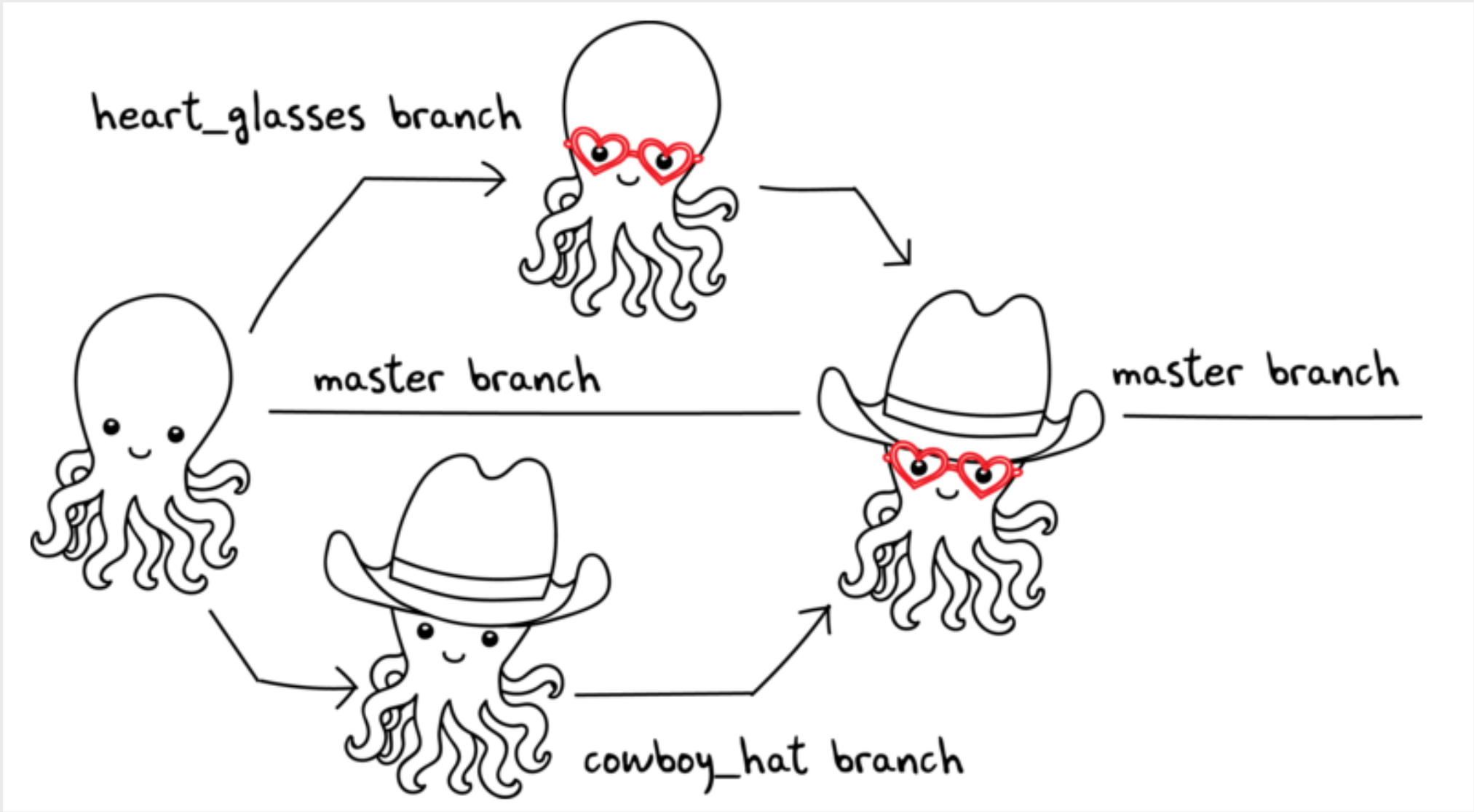


GISc WORKFLOW

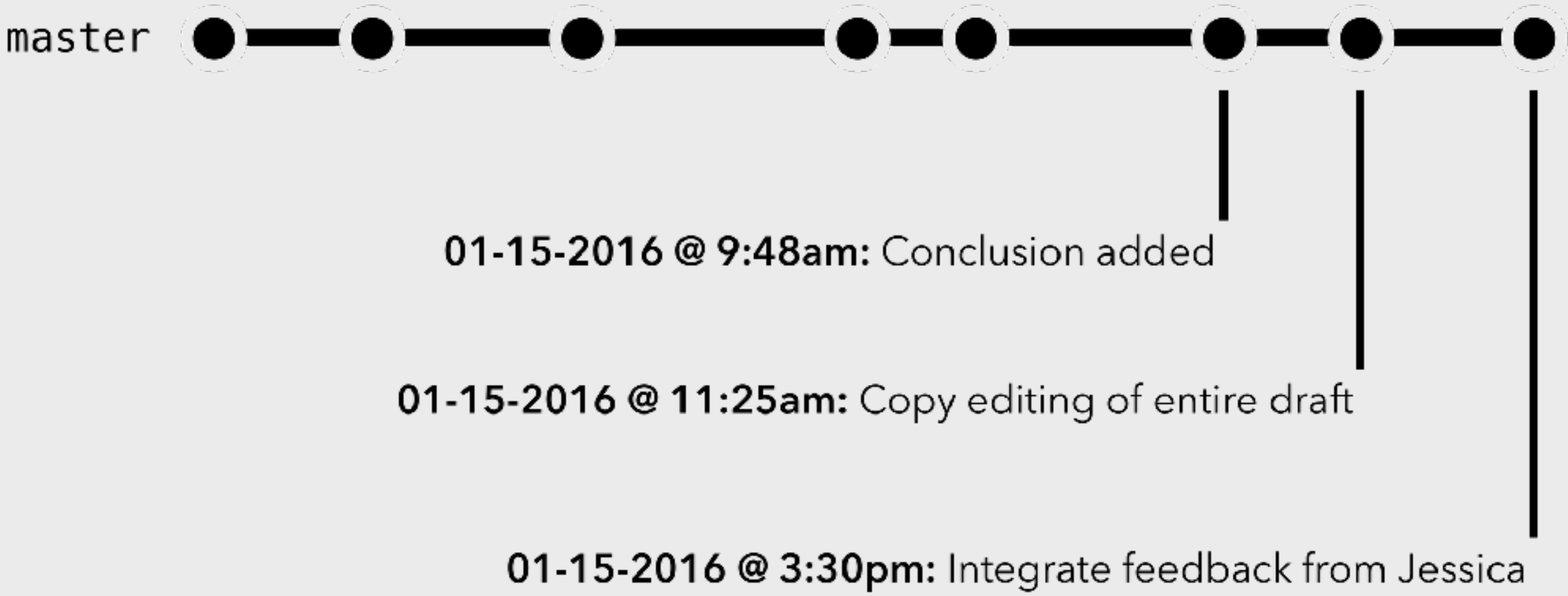


2 ADVANCED GITHUB

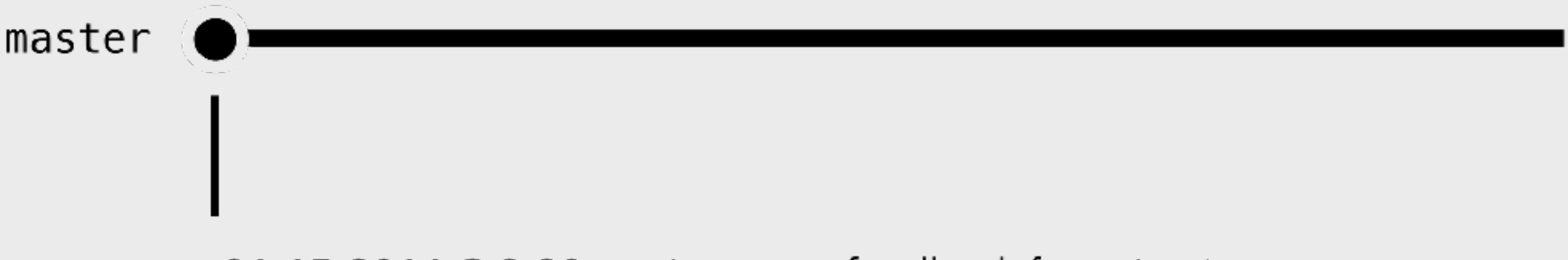
BRANCHING & THE FINAL PROJECT



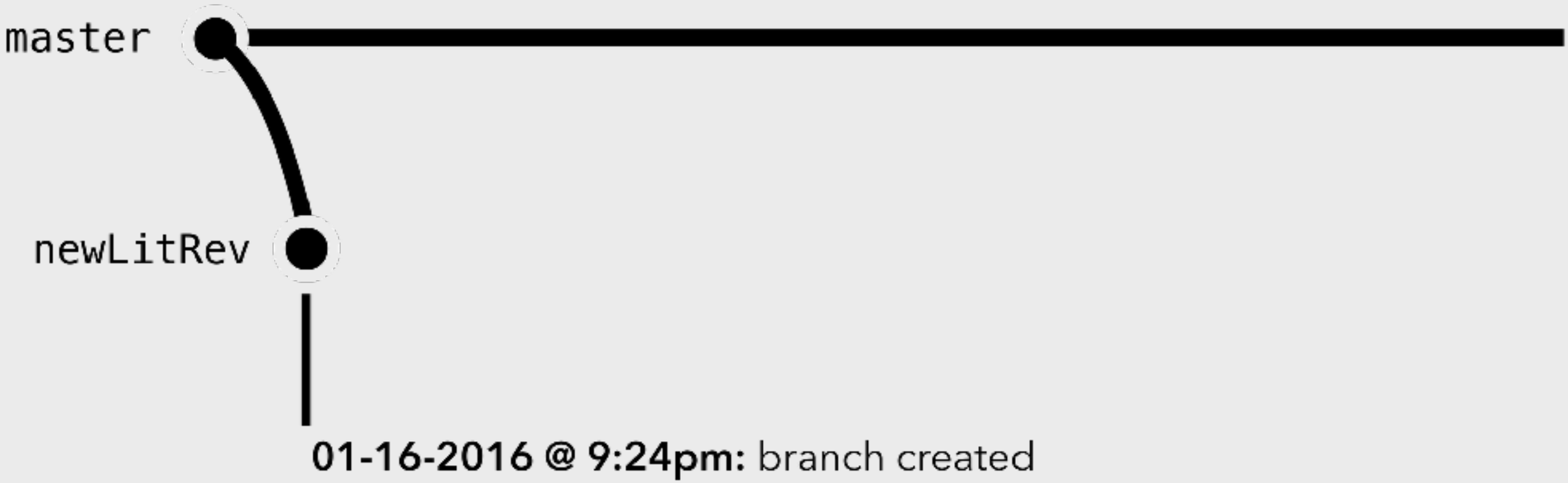
BRANCHING


master
01-15-2016 @ 9:48am: Conclusion added
01-15-2016 @ 11:25am: Copy editing of entire draft
01-15-2016 @ 3:30pm: Integrate feedback from Jessica

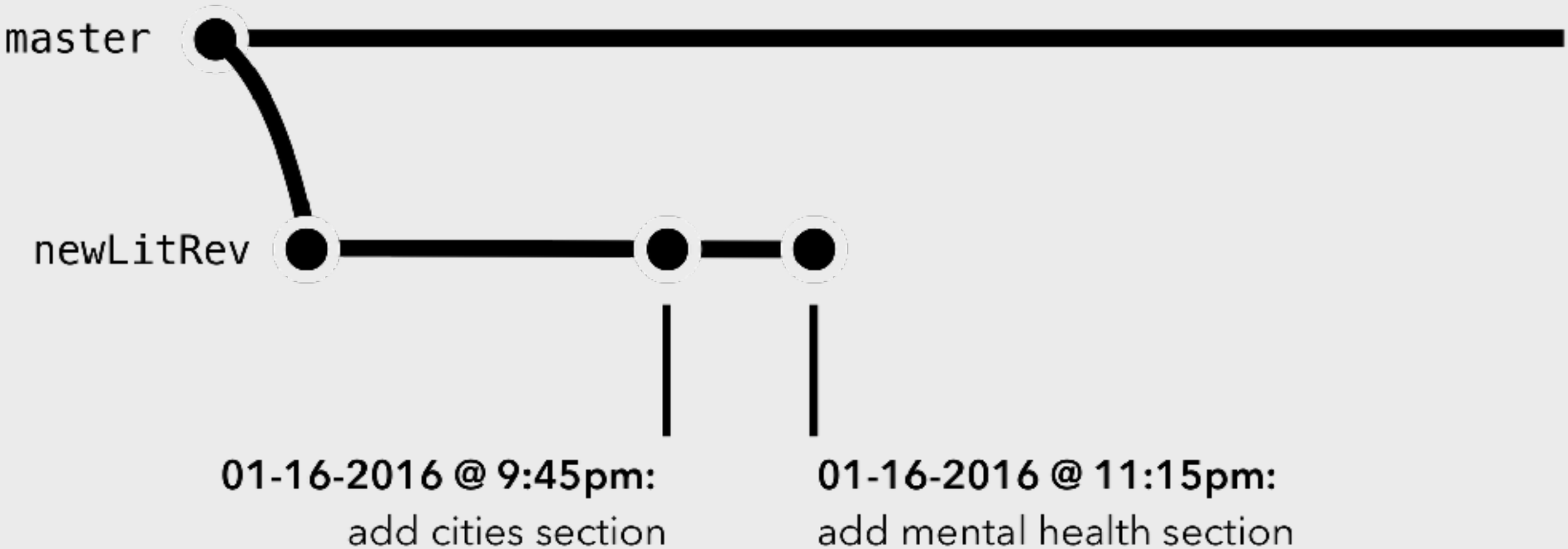
BRANCHING



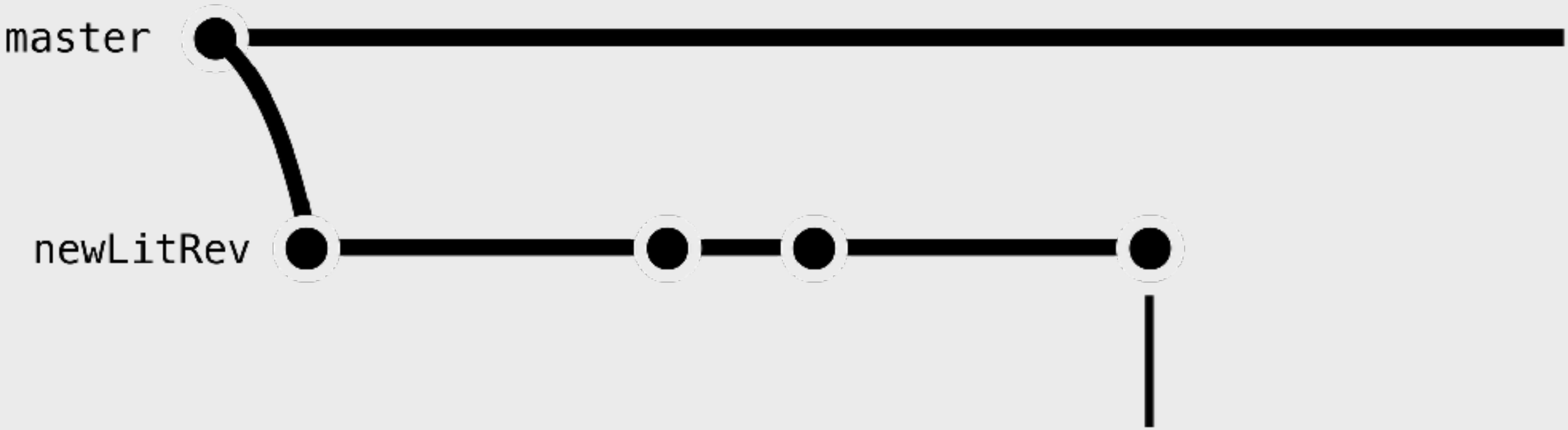
BRANCHING



BRANCHING


master
newLitRev
01-16-2016 @ 9:45pm:
add cities section
01-16-2016 @ 11:15pm:
add mental health section

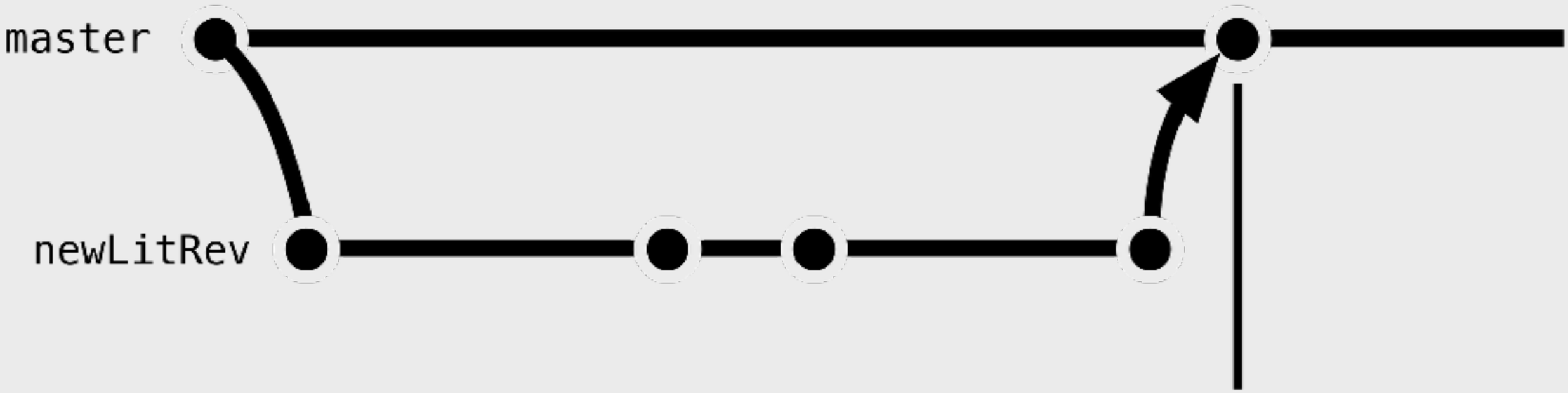
BRANCHING



01-17-2016 @ 9:10am: edited lit review

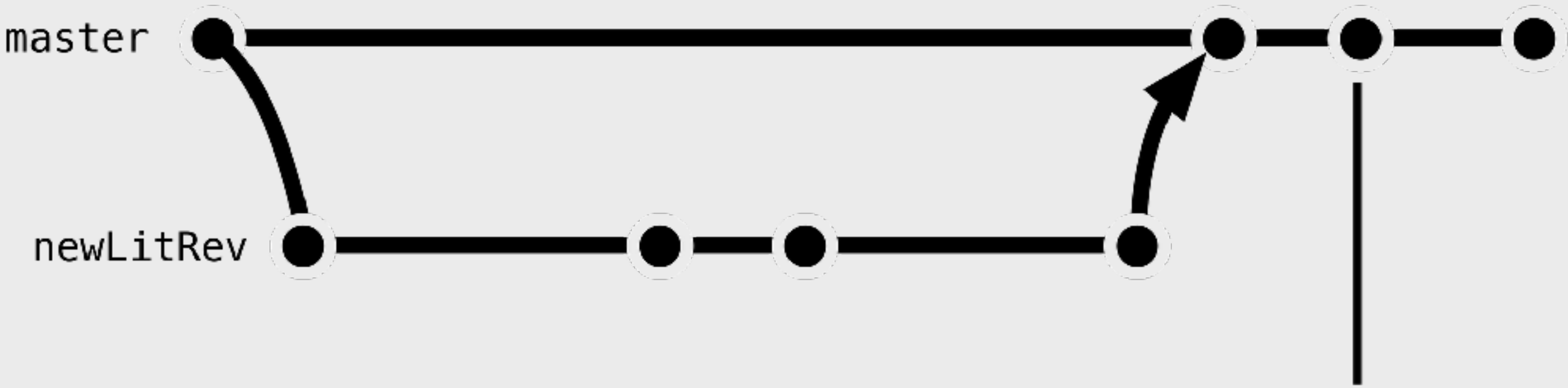
01-17-2016 @ 9:25am: open pull request

BRANCHING



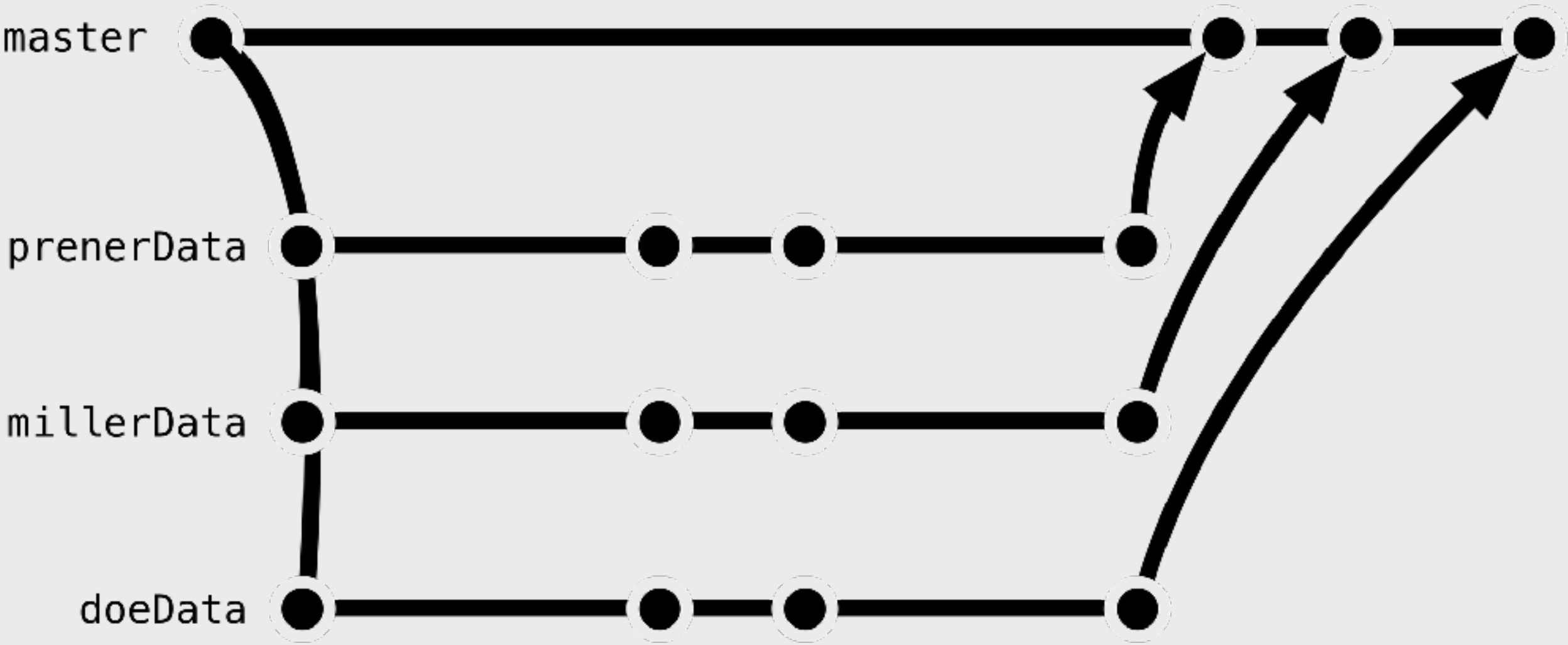
01-17-2016 @ 1:30pm: accept pull request, “delete” branch

BRANCHING



01-17-2016 @ 2:22pm: edit transition into new lit review

BRANCHING



3 CREATING NEW VARIABLES

GOALS – 'TIDY' DATA

- "Each variable forms a column" with a unique, simple, intuitive name
- "Each observation forms a row" with no duplicates
- "Each type of observational unit forms a table"

- In GISc, it is important to understand if duplicates are truly duplicates, how that will impact mapping
- Values should be clearly defined and consistently applied
- Datasets should be subset if possible to make management easier and to condense the size of shapefiles down the road

see Wickham 2014

TWO STRATEGIES

1. "Copy and Recode" - create a duplicate version of the variable and then rearrange the values. This works well for continuous variables that are being recoded into ordinal or binary variables.

```
. generate newVar = oldVar

. recode newVar (0/10 = 1) (11/20 = 2) (21/30 = 3)
```

TWO STRATEGIES

2. "Create and Replace" - create a new, empty variable and then replace it with specific values based on another variable. This works well for re-ordering categorical variables and working with string variables.

```
. generate newVar = .
. replace newVar = 1 if oldVar < 11
. replace newVar = 2 if oldVar >= 11 & oldVar < 21
. replace newVar = 3 if oldVar >= 21 & oldVar < 31
```

TWO STRATEGIES

2. “Create and Replace” - create a new, empty variable and then replace it with specific values based on another variable. This works well for re-ordering categorical variables and working with string variables.

```
. generate str newVar = “”
. replace newVar = “no” if oldVar == “foo”
. replace newVar = “yes” if oldVar == “bar”
```

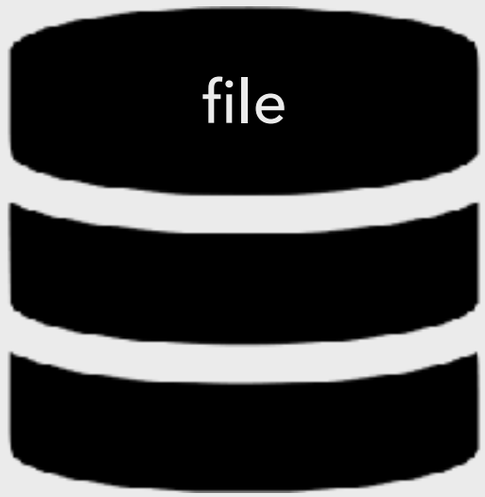
TWO STRATEGIES

2. “Create and Replace” - create a new, empty variable and then replace it with specific values based on another variable. This works well for re-ordering categorical variables and working with string variables.

```
. generate str newVar = “”
. replace newVar = “no” if strpos(oldVar, “foo”)
. replace newVar = “yes” if strpos(oldVar, “bar”)
```

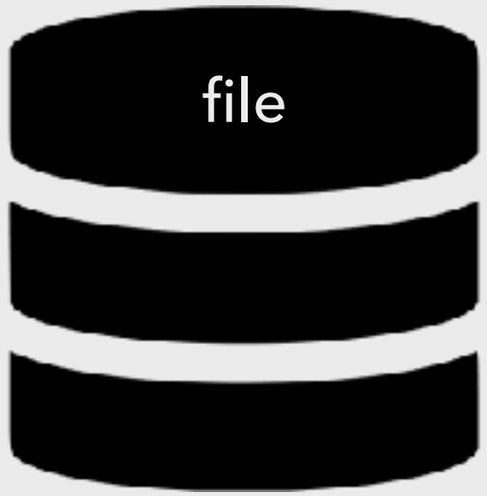
4 GEODATABASES

DATA STORAGE OPTIONS



DATA STORAGE OPTIONS

Ideal for storing and sharing single files, but can be inefficient and difficult to store.



Ideal for storing large amounts of spatial data from both organizational and efficiency approaches. Can get large quickly, which may have ramifications.