

# *SERPM 8.0 Model Usability Plan*

RTTAC-MS Approved – 03/15/2017

# *Model Usability Topics*

---

- Implementation
- Installation
- Structure
- Scenario Development and Configuration
- Interface
- Run Modes
- Reports and Logging
- Document Structure:
  - » Proposed Usability Improvements
    - Current Approach
    - Challenge it Presents
    - Proposed Enhancement

# Implementation

## Current / Challenge

- Current:
  - » Cube 6.4 with cluster
  - » CT-RAMP
  - » R statistical software
- Challenge:
  - » High-performance computing and memory requirement (144GB)
  - » Requires Cube Cluster additional license

## Proposed Enhancement

- Maintain implemented software, but support 'light-weight' operations
  - » Primary deployment: single workstation
  - » Run modes to include 'assignment-only' allowing operation on a system without 144GB RAM
    - Option to run assignment only without cluster
  - » Continue to support multi-workstation (utilized by CT-RAMP process only)
- Investigating cloud deployments (separate memo)

# Installation Process

---

## Current / Challenge

- CT-RAMP requires multi-step installation process
  - » Install Java
  - » Set Windows variables
  - » Copy executables and setup DLLs
  - » Configuration of Java properties files
- R statistical software require packages for analysis
  - » Could include R Studio for custom analysis

## Proposed Enhancement

- Installation 'wizard' to guide the user through the installation process of non-Cube elements
  - » Will setup single-workstation with Java
  - » Other setups (multi-workstation) will require manual configuration
- Include unit tests to verify correct installation of ABM components

# Upgrade Process

---

## Current / Challenge

- New versions require replacing entire model folder

## Proposed Enhancement

- Installation 'wizard' will always install latest version of the model
- Upgrade 'patches' will be developed for limited-updates allowing users to apply the patch rather than re-install the entire model

# Model Structure

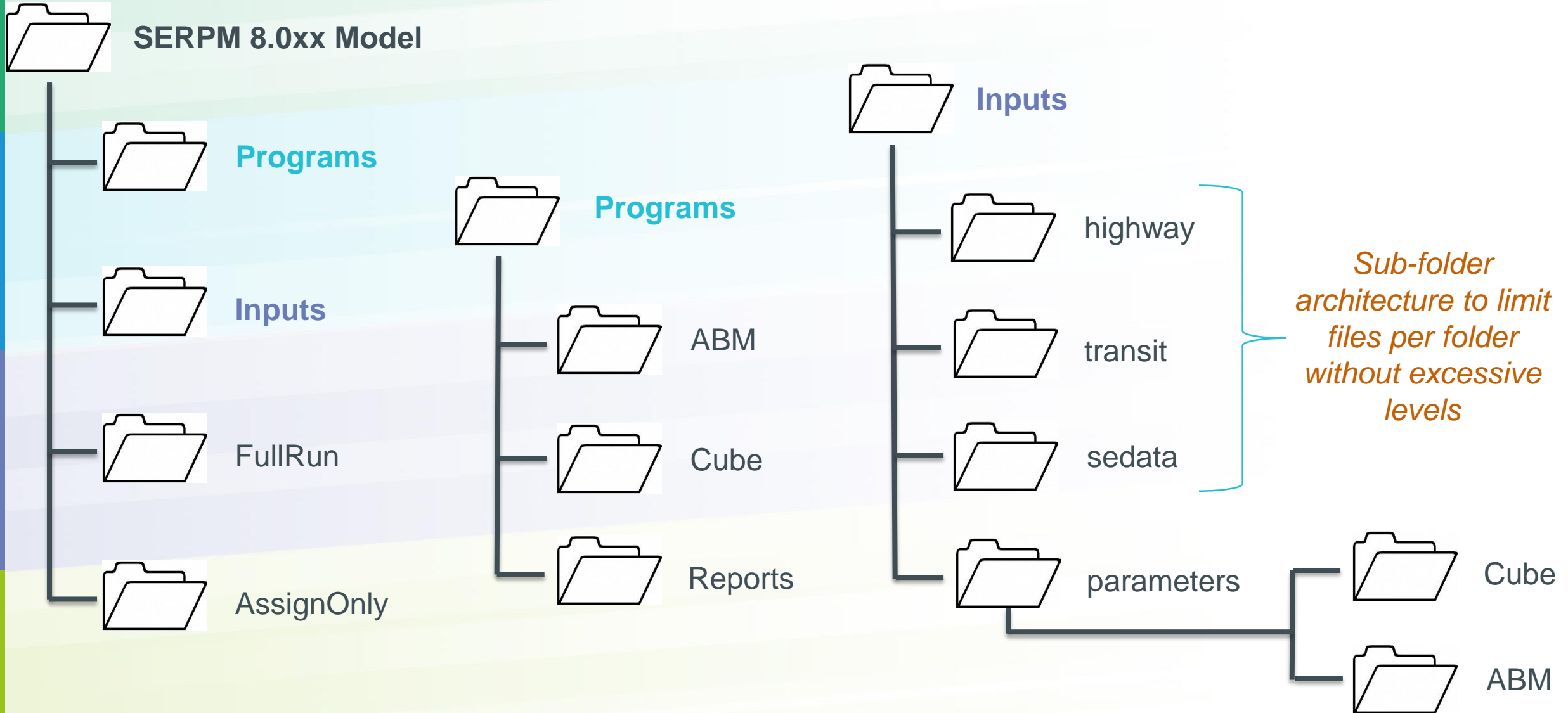
## Current / Challenge

- SERPM7 must be located in C:\SERPM7 folder
  - » Requires computer configuration with ample space on C: drive (i.e., not a windows drive / data drive configuration)
  - » Limits flexibility to have multiple model versions on same system
- Scenario management requires input folders to be created manually and synchronized through the Cube Scenario Manager
  - » Supports need to be able to copy inputs from a scenario for archiving
  - » Input folders are not updated from model run
- Parameter / Configuration files are stored in multiple places
  - » Input folder
  - » Ctramp folder

## Proposed Enhancement

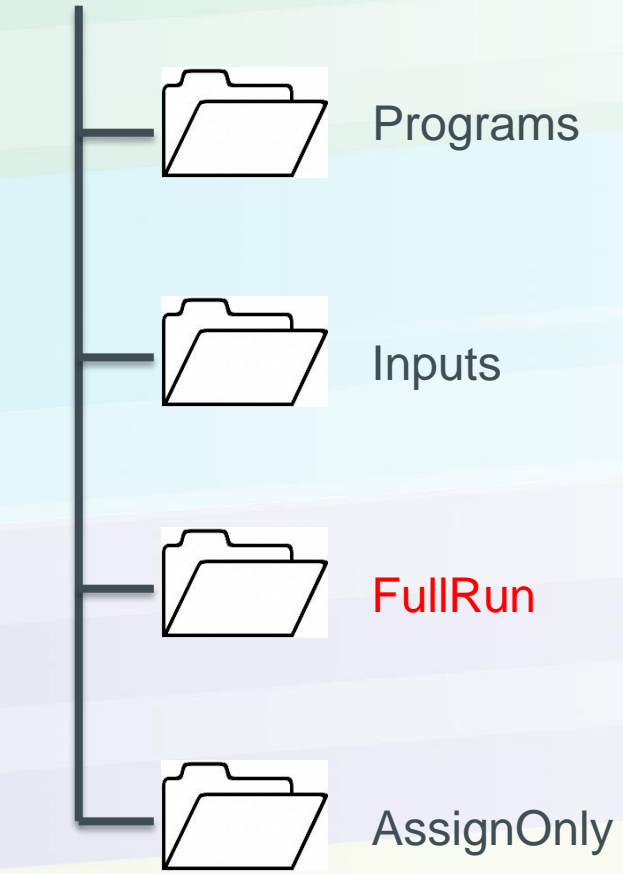
- Any 'reasonable' installation folder supported
  - » Will parameterize as necessary the path to programs, input files
  - » 'Reasonable' to be determined by Cube (e.g., no '@' in path)
- Leverage Cube Catalog to organize scenarios
  - » Single Input folder
  - » Hierarchical Output folders
  - » Common files (programs) stored in root directory
- Create an inputs folder in Outputs to maintain archive
  - » Copy of selected inputs
  - » Must be clear to user that edits need to be made in main Inputs folder

# Model Structure: Folders

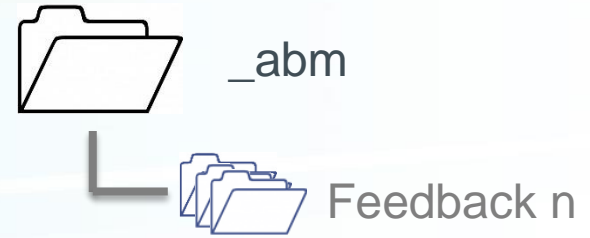


 **SERPMP 8.0xx Model**

 **FullRun**



 **R2015**



 **R2045**



# Model Files

## Current / Challenge

- File names do not follow FSUTMS standard
  - » \*\_yya.\* input / \*\_ayy.\* output
  - » Standards (Data Dictionary 12/2005) do not cover ABM
  - » \*.MAT resolves to “Microsoft Access Table Shortcut”
- Files from intermediate Cube programs are not removed, increasing the output folder size
- Previous iterations of speed feedback are overwritten, preventing analysis of model convergence.

## Proposed Enhancement

- As applicable, implement FSUTMS naming and attribute guidelines
- Optionally set \*.MAT to default to Cube during model installation
- Use ‘~’ prefix and optionally delete intermediate Cube program files after each iteration/run
- Speed feedback outputs – optionally copied to a subfolder or removed at the start of the next iteration

# *Scenario Development and Configuration*

---

- Highway and Transit Network Development
- Input Data Development
  - » Synthetic Population
  - » Employment
- Input Data Verification
- Scenario Configuration Definition
- Scenario Configuration Verification

# Scenario: Highway/Transit Network Development

## Current / Challenge

- Highway network built in GIS with projection and 'multi-link curvature'
- GeoDatabase worked-around in the model process
  - » Requires ArcGIS license to run model
  - » SERPM 7 user experience has been to work with \*.NET files directly and not work within the GeoDB
- Model year and build/no-build scenarios are implemented through separate network files

## Proposed Enhancement

- GeoDB
  - » Will investigate removing from model process and supporting TAP process through individual dbf files
  - » Can be useful for mapping
    - Reporting GeoDatabase will be maintained with zonal-shape files included as layers
- Scenario networks
  - » Maintain separate networks (no master network)
- Model inputs will be \*.NET file for highway and line file for transit (i.e., not GeoDB)

# Scenario: Synthetic Population Development

## Current / Challenge

- Synthetic population produced by PopSynIII based on TAZ marginal control totals
  - » Requires SQL Server to run
- Synthesizing changes in a single zone requires re-running the entire PopSynIII and may result in population changes in other zones
  - » Complicates site-analysis projects where the only change in the model should be limited to a small number of zones
  - » SERPM7 includes an R-script process to update individual zones between two full PopSyn runs

## Proposed Enhancement

- Utility to populate specified zones according to targets set by user, for example:
  - » Workers by HH
  - » Income by HH
- Set weights from ACS sample according to these inputs and sample

Single Zone Synthesizer		zone	<input type="text" value="1234"/>		
		households	<input type="text" value="5,000"/>		
size1	<input type="text" value=".4"/>	inc1	<input type="text" value=".05"/>	wrk0	<input type="text" value=".2"/>
size2	<input type="text" value=".2"/>	inc2	<input type="text" value=".05"/>	wrk1	<input type="text" value=".5"/>
size3	<input type="text" value=".2"/>	inc3	<input type="text" value=".2"/>	wkr2	<input type="text" value=".2"/>
size4	<input type="text" value=".2"/>	inc4	<input type="text" value=".3"/>	wrk3	<input type="text" value=".1"/>
		inc5	<input type="text" value=".4"/>		

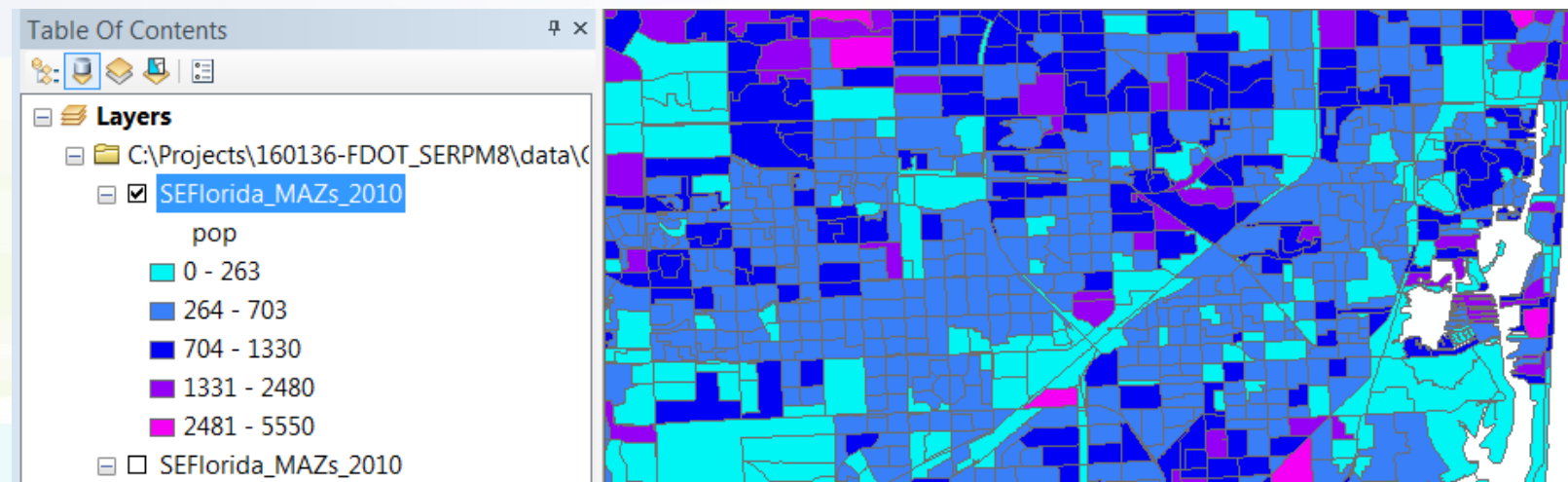
# Scenario: Employment Data Development

## Current / Challenge

- Employment data defined in CSV data file
- Changes to an MAZ involves
  - » Identifying the MAZ through GIS
  - » Finding the row in the MAZ data file
  - » Making changes and saving to a new file name
  - » Updating the scenario manager
  - » Updating the serpm\_abm.properties file

## Proposed Enhancement

- Attach MAZ data to shape-file, allowing editing to be done through GIS programs
  - » Configured GIS maps (\*.MXD) files saved to facilitate loading/editing through Cube, ArcGIS or other GIS program
  - » Populate map file with layer package showing summaries of common attributes (employment, population, etc.)
- MAZ input file to be the DBF file associated with the shape-file
  - » Defined in Cube Scenario Manager and passed down to ABM serpm\_abm.properties file



# Scenario: Input Data Verification

---

## Current / Challenge

- Issues with input files cause crashes later in model
  - » Once the modeler has stopped paying attention
- Or, worse yet, are unnoticed: non-catastrophic, but important errors in
  - » Network coding
  - » Socioeconomic data

## Proposed Enhancement

- Run-Time Checks
  - » Built into Cube Catalog
  - » Run on first iteration – log descriptive error / warning
- Networks
  - » Highway skims have valid times, distances, speeds between interchanges
  - » Transit lines can be loaded on the Highway network
- Socioeconomic inputs
  - » Zones with zero households, workers, children
  - » Total employment <> sum of parts
  - » Alignment of supply / demand
  - » Workers to employment
  - » Children to enrollment
  - » College students to university enrollment

# Scenario: Definition and Verification

## Current / Challenge

- Scenarios defined using Cube Scenario Manager and serpm\_abm.properties file
  - » Scenario manager includes many parameters that the average user would not change. E.g.,
    - Number of zones
    - Walking speed
    - Capacity factors
    - PCEs
  - » Scenario manager does not specify which serpm\_abm.properties file is being used
- After model run, not obvious what model parameters were used.

## Proposed Enhancement

- Cube Scenario Manager
  - » Present inputs that are most frequently changed: highway/transit networks; synthetic population; SE data files; special generators
  - » Catalog control: number of iterations
  - » Specifies next level configuration files (global and abm)
- Global configuration file
  - » Flat text file readable by Cube and CT-RAMP
  - » Contain common settings that will not be changed often (zones, relative gap, auto operating costs, seed skins)
- ABM configuration files
  - » Specific parameters and model definitions for ABM components
  - » UECs, visitor rates, etc.
- Configuration report (next slide)



# Scenario Configuration Report

## Record scenario parameters

- Global configuration file and ABM parameter files copied to output folder

## User verification

- Upon model start – compare model inputs against default for each scenario
- Highlight changes in:
  - Cube Scenario Keys
  - Input file dates and sizes
- Optionally prompt user to verify and accept changes

## Mock Up of Report

	Default Scenario			Current Scenario		
	Filename	Size	TimeStamp	Filename	Size	Date
Highway Network	S8_15A.NET	1.02MB	July 1, 2017, 10:00:00AM	same	same	same
Transit Line File	TROUTE_R15.LIN	256KB	July 1, 2017, 10:00:00AM	TROUTE_R15_Test.LIN	257KB	August 15, 2017, 10:00:00AM
Housholds	households.csv	1.02MB	July 1, 2017, 10:00:00AM	same	same	same
Persons	persons.csv	2.02MB	July 1, 2017, 10:00:00AM	same	same	August 15, 2017, 10:00:00AM
MA7 File	maz_data.csv	543KB	July 1, 2017, 10:00:00AM	same	same	same



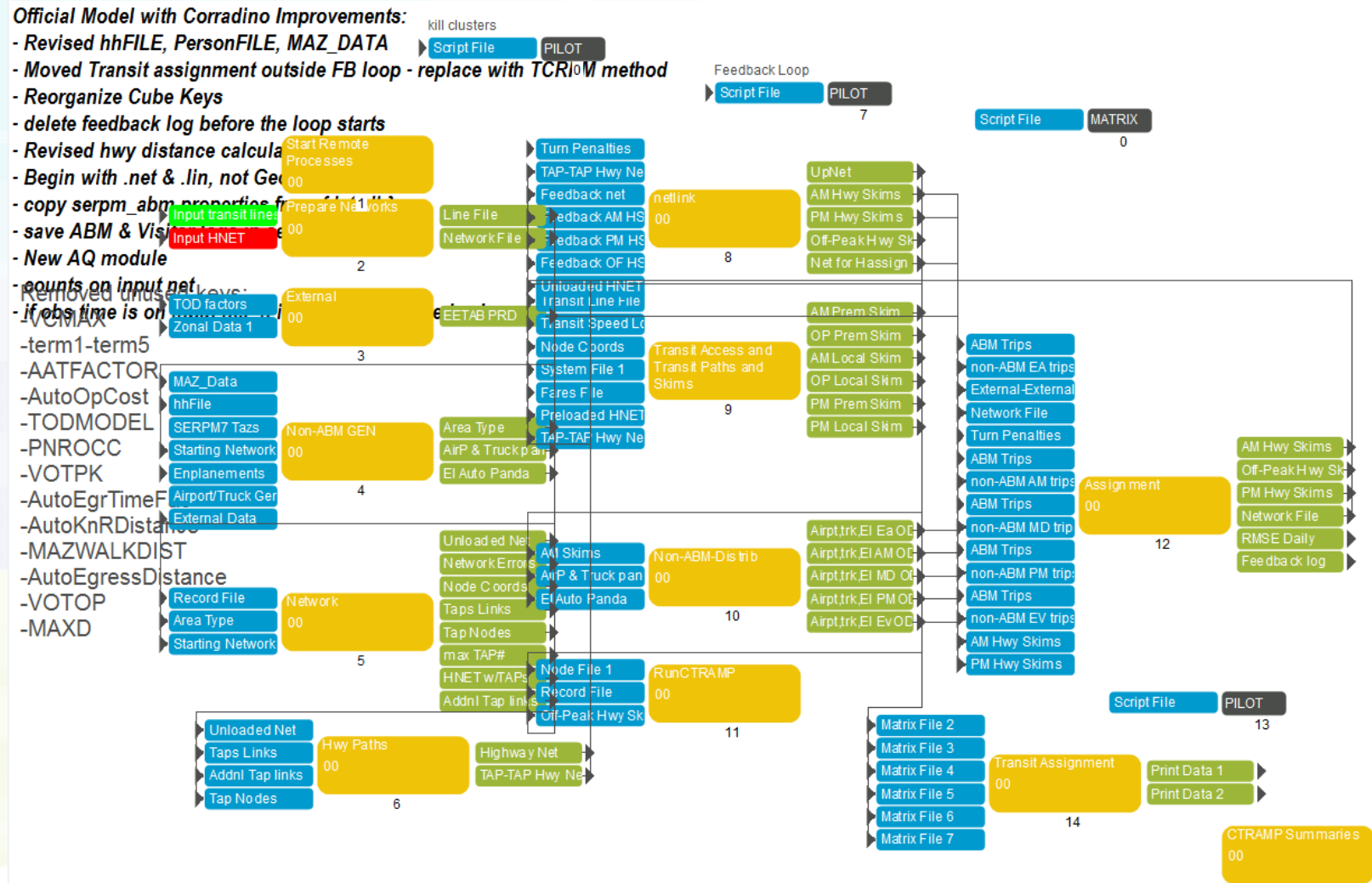
# Model Interface: Current / Challenge

## Cube Catalog

- Use of Cube's input/output boxes creates overlap with different display resolutions
- Text descriptions are hard to read

## Utilities

- Optional programs inserted in various places in the Catalog



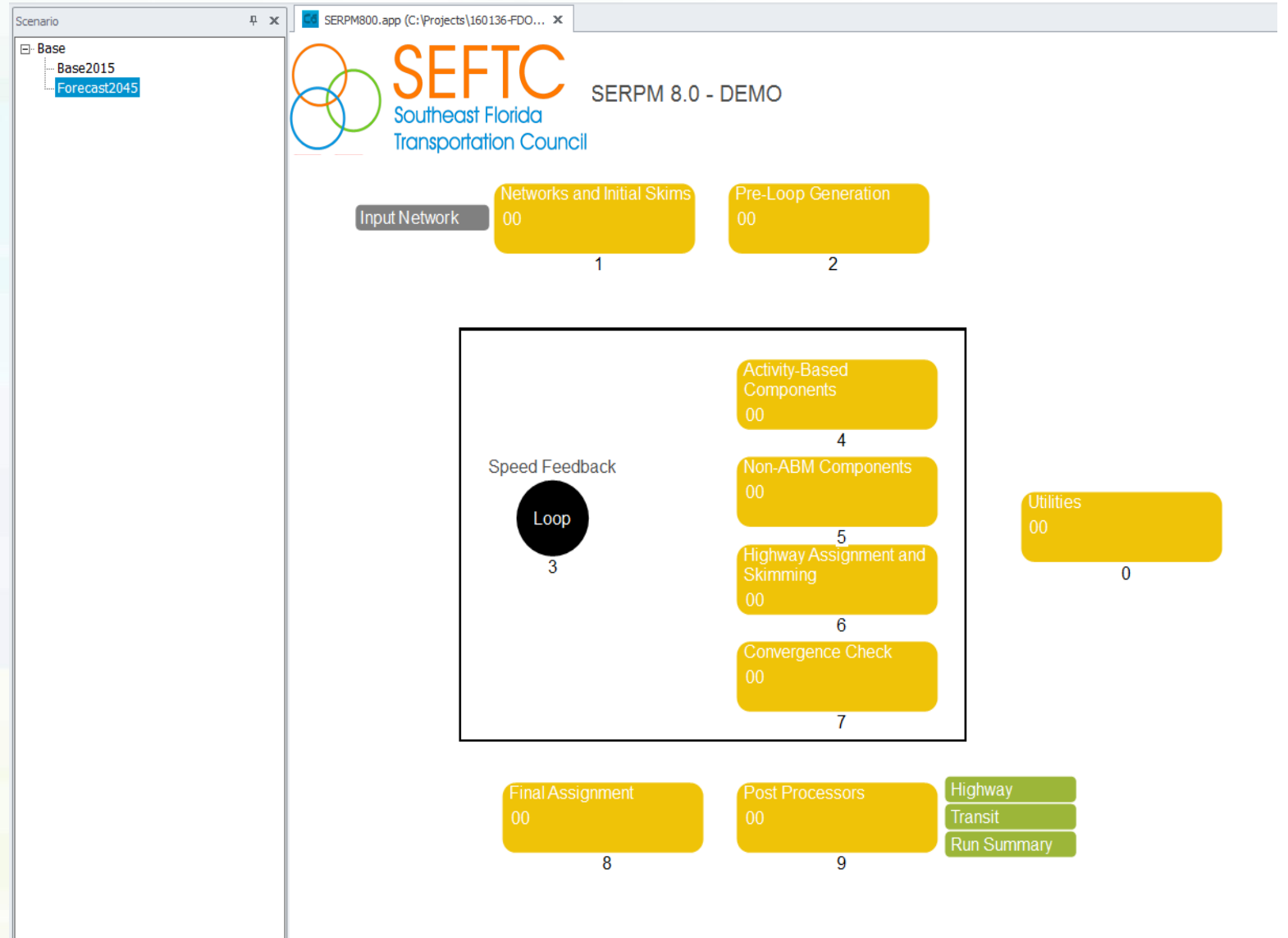
# Model Interface: Proposed Enhancement

## Cube Catalog

- Implement hierarchical structure to simplify high-level
- Compartmentalize operations
- Reduce publicizing of input/outputs to few reports

## Segment Utilities

- Optional programs placed in separate application group



# Model Interface: Scenario Definitions

- Simplify scenario keys
- Organize scenarios by hierarchy

The screenshot displays the SERPM 8.0 Model interface. On the left, a tree view under 'Scenario' shows a hierarchy: 'Full Run' contains 'Base2015' and 'Forecast2045'. 'Base2015' contains 'TestScen1' (highlighted in blue) and 'TestScen2'. 'Forecast2045' contains 'TestScen3' and 'TestScen4'. The main panel is titled 'SERPM 8.0 Model - DEMO' and lists several data sources: Highway Network, Transit Network, TAZ Data, MAZ Data, Persons File, and Households file. To the right of these labels are six empty rectangular input fields.

# Model Interface: CT-RAMP

---

## Current / Challenge

- CT-RAMP called as a separate program from Cube
  - » Errors from CT-RAMP process are not detected by Cube Process
  - » Cube continues running until trip tables are not found
- Random crash occasionally occurs requiring restart of the model
  - » Can restart at current step
  - » Lost computational time waiting for the modeler to detect, identify error, and restart model

## Proposed Enhancement

- Explicitly trap errors from CT-RAMP
- Optionally restart step automatically if 'random-crash' occurs

# Run Modes

## Current / Challenge

- Full model run
  - » Specify starting index of feedback loop
- Assignment only
  - » Launched as an 'Application Group' run that requires users to have populated the output folder with necessary files
  - » Running Highway and Transit is a 2-step process
- Windowed-Area Model
  - » Geographic sampling
  - » Sub-area highway evaluation summaries

## Proposed Enhancement

- Full Run
  - » Continue implementation with user specified index
- Highway and/or Transit Assignment only
  - » Defined in separate catalog file
  - » Scenario key that points to location of trip tables for assignment
- Windowed-Area Model
  - » Details to be discussed in next project status meeting

# Reports and Logging

## Current / Challenge

- ABM reports
  - » SQL Server: not in common use anymore
  - » Optional cube summaries: difficult to modify
  - » Logs written back to program location
- Cube outputs
  - » Loaded networks by time period and daily summary networks – no default VPR files to aid visualization
  - » Other outputs across multiple files
  - » Logs (\*.PRN files) written to same folder as outputs

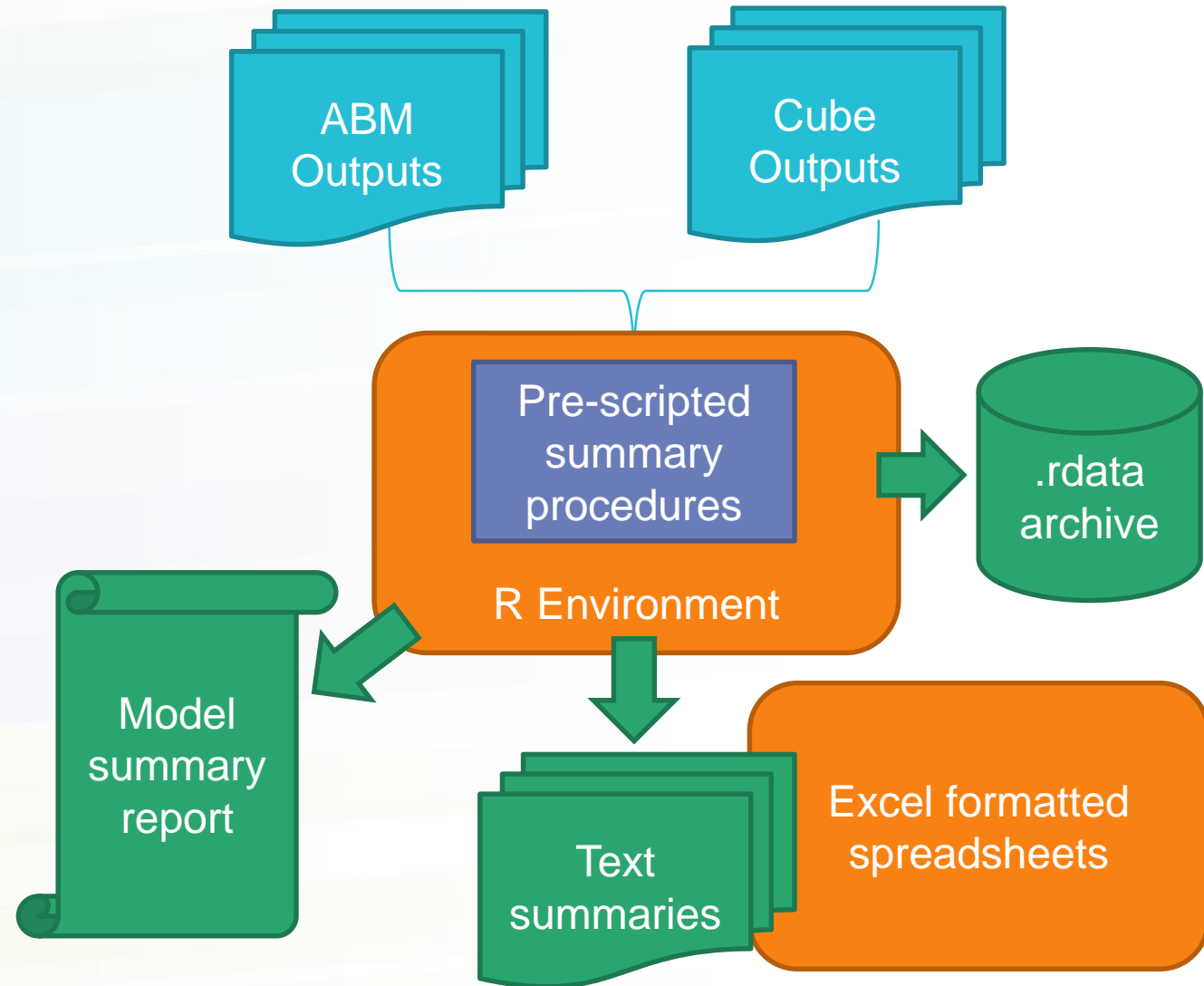
## Proposed Enhancement

- Leverage R statistical software to process non-graphical output data
- Develop standard graphical formats
- Consolidate logs into single location
- Specifics described in more detail in following slides

# Reports and Logging

## Reporting process

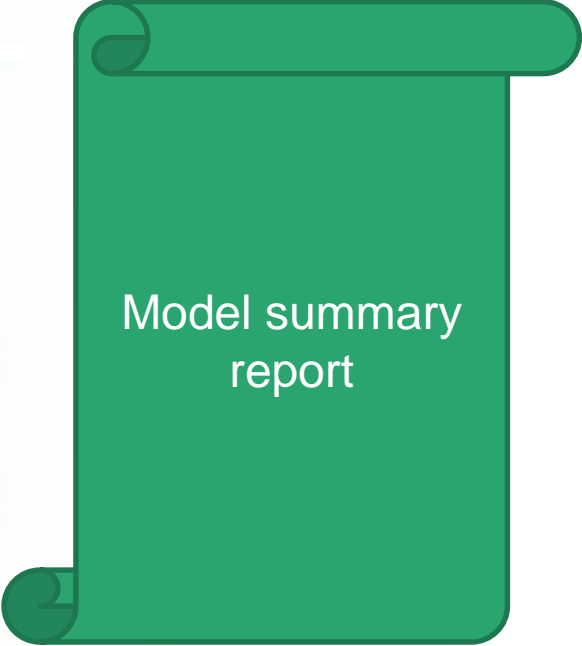
- » Summarize data using R
  - Export summaries to excel for manipulation
  - Save .rdata file for further analysis
- » Run summary – html or pdf
- » Detailed reports in Excel



# Run Summary

---

- Input file / parameters with changes highlighted
- Aggregate level statistics
  - » Population / Households / Employment
  - » Trips:
    - Total person, vehicle, mode, district/county
    - Rates
    - Shares by mode



Model summary  
report





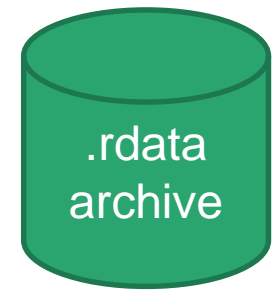
➤ Prepared with R, viewable with Excel

» Ok to assume Excel availability?

➤ Option to select multiple scenarios and compare – e.g., AV Scenario sheets

➤ Reformat existing reports (SERPM\_REPORTS.XLS)

	County	Income	Person Type	Occupation	Tour Purpose	Trip Purpose	Mode
Persons	x		x				
Transponders	x						
District - district tours					x		
Number of tours	x		x		x		x
Number of trips	x		x			x	x
Tour distance / time			x		x		
Time Spent Traveling		x	x				
CBD Trips				x			



R environment saved file

Loading file brings up all data used to create reports

Custom summaries, plots, other exploratory analysis can be done through RStudio

The screenshot displays the RStudio environment. The main window shows a data table with columns: hh\_id, person\_id, person\_num, age, gender, type, value\_of\_time, activity\_pattern, imf\_choice, and inm. The table contains 18 rows of data. The Environment pane on the right shows the Global Environment with Data objects: Ind\_trip (6189941 obs. of 18 variables), per\_trip (6189941 obs. of 30 variables), and person (1750027 obs. of 13 variables). The Functions pane shows a function named SERPM\_Summar... function(). The Files pane shows a histogram titled "Histogram of Ind\_trip\$stop\_period". The Console pane shows the following commands and output:

```
5: Eating Out 9005
6: Maintenance 8454
7: shop 17673
> view(Ind_trip)
> view(person)
> hist(Ind_trip$stop_period)
>
```

# *Standard Reports - Graphical*

---

- Leverage Cube to display networks
- Loaded network summaries saved in report folder
  - » VPR file associated with input and output networks
    - Identify facility types
    - Volumes by Time of Day
    - Comparison to count data
    - Volumes by vehicle segment
    - Others as needed

# Logging

---

- Cube generates PRN files per program
- ABM logfiles
- All log files to be stored in \_logs folder
  - » Central location to simplify troubleshooting and facilitate users sending information back to developers
- Text parsing script to identify issues
  - » Find the ErrorLevel 2 across Cube Cluster files
  - » Identify errors from ABM logs