

So You Want a Data Mining/Warehousing Package: Starting the Process and Beyond

Dr. Robert Springer
Executive Director of Institutional Effectiveness
Office of Institutional Research and Assessment
Elon University

Contextual Information...

* **Elon University**

- * Location: Central NC
- * Private, non-affiliated
- * ~ 6,500 total students
- * ~ 5,800 undergrads
- * 12:1 student-faculty ratio
- * 6 Graduate Programs
 - * DPT, PAS, JD, MBA, MED, Interactive Media
- * 97%-98% fulltime students
- * 62% live on campus
- * ~ 89% faculty are PhD
- * ~91% 1-yr retention rate
- * ~77% 4-yr graduation rate
- * ~81% 6-yr graduation rate
- * Endowment <\$200M

...Contextual Information

- * Office of IR & Assessment = 2014
- * 3 person operation
 - * Director
 - * Associate Director
 - * Assessment Specialist

5-Year Plan

1. Mission Statement
2. Assessment Plan
3. Technology Plan
4. Reporting Plan
5. Professional Develop Plan

Target Audience

- * Academic Affairs and other academic related offices
 - * Provost, Deans, Chairs
 - * Registrar's Office
 - * OIR&A
- * Other offices/areas will be considered as successful rollout continues
- * **Basic Information – baby steps**

Current VA Process

Extract-Transfer-Load (E-T-L)

- * (E) Extract data from Colleague
- * (T) Run through SAS to create a SAS dataset
 - * Maintenance occurs - clean up, organize, labeling, etc...
- * (L) FTP to the cloud (Pinnacle Solutions)
- * (L) Remote Desk Top to move into SAS-VA laser locations
- * (L) In SAS-VA Upload and register the table (dataset)

Long Term Goal

- * Move data/tables directly in VA
 - * Developed a data-pull plan
 - * Pull census data for current term (pre) and most recent past term (post) on same day.
 - * Census = ~ the 10th day of classes normal term
 - * Census = ~ the 3rd day of classes Winter and Summer terms
 - * This will allow for official reporting (pre), final grades, enrollment, study abroad transfer credit, and other “final” (post) counts to occur.

Helpful SAS Code

Appending Files

```
ods html close; /* close previous */  
ods html; /* open new */  
title ' ';
```

```
libname Pinnacle 'C:\Users\springer\Desktop\Pinnacle\';
```

```
data majors;
```

```
proc append base=pinnacle.majors
```

```
data=pinnacle.enr2012major;
```

```
*data=pinnacle.enr2013major;
```

Notes: use FORCE (or all variables are same length)

Helpful SAS Code

Single to Multiple Records

Example: Double/Triple Majors (or minors)

```
array major123[3] $ major1 major2 major3;
```

```
do i = 1 to 3;
```

```
major = major123[i];
```

```
if major ne ' ' then output;
```

```
End;
```

```
Run;
```

Helpful SAS Code

Multiple Files – Single Folder

```
libname Pinnacle 'C:\Users\springer\Desktop\Pinnacle';  
run;
```

```
ods html close; /* close previous */  
ods html; /* open new */  
title ' ';
```

```
data pinnacle.AllCourseSections (drop=newtitle sl elr lab);  
infile 'C:\Users\springer\Documents\OldCDocs\DAT\crse\  
Crse2014-15\*.CSV' dlm=', ' pad DSD missover firstobs=2;
```

Labeling Hints

- * SAS will alpha sort...
 - * First-year, Junior, Senior, Sophomore
 - * Fall, Spring, Summer, Winter

Spaces are considered a lesser value (^ = space)

^^^First-year

^^Sophomore

^Junior

Senior

^^^Fall

^^Winter

^Spring

Summer

Other Hints

- * What you create in a report – stays in that report
- * There is no undo button
- * If you import versus FTP files and a re-set has to occur to the server, that imported file must be re-imported... meaning you must do the re-import
- * Measures can be duplicated and can become a Category
- * Categories can be duplicated and can NOT become a Measure

Other Hints

- * If joining files, you may want to code variables in such a way that you know what table/dataset the variable came from.
 - * You may have credit hours, gender, race/ethnicity,... residing in different tables and need to join those tables
- * Think about where you want to do your grouping/aggregates
 - * Undergraduates/Graduates
 - * Colleges/Schools/Divisions

Data Dictionary

- * Do you have one?
- * Chances are you've been using some standards already in your reports.
- * With greater outreach, expect more questions
 - * More questions will help you structure your data for better and more meaningful use

Finding New Issues

- * You will see things you might have never known in your data
 - * e.g., Graduate students with a minor or a second degree
- * Be prepared to lead the discussions on error/log reports and additional standards/processes