Paper 195-2012

SAS® and Microsoft SharePoint: Project and Resource Management Tools and Techniques for Large, Globally Distributed Programming Teams

W. Stetson Line, Amgen Inc., Thousand Oaks, CA

ABSTRACT

Most large companies employ a variety of in-house, off-shore, and service provider resources. Staff members often work in a variety of time-zones and can have different vacation and holiday calendars based on local, ethnic, and religious customs. How can managers effectively plan for resource demand peaks and valleys in such an environment?   This paper presents a modest and scalable approach using SAS and MS SharePoint that provides a useful framework for both long-term forecasting and short-term load-balancing decisions in the management of globally distributed teams.

INTRODUCTION

Managers of international teams face a number of challenges: Planning for work peaks and troughs is complicated as staff locale time-zone and time-off diversity is not handled by common resourcing tools. The tasks and roles of remote workers may need to be limited as a consequence of not having enough direct communication with functional partners operating in different time zones. Job performance must be guided and gauged without the intimate involvement of the remote manager. All staff members need to provide updates pertaining to their work schedules and the status of their deliverables in a way that is timely and actionable or else "in-box" problems can become serious. These challenges and others require innovative processes and tools.

Over the last year a solution has been developed using SAS, MS SharePoint, and Java/Groovy that has proven useful in addressing the above-stated challenges.   This solution also functions as a communications hub for distributed staff and provides a means of work monitoring and tactical assistance in the performance of daily tasks.

Our CDISC Consultancy and Implementation (CCI) team is part of the Global Statistical Programming department of Amgen, Inc. A core CCI mission is to assist clinical study teams in the specification, production, and quality assurance of CDISC Study Data Tabulation Model (SDTM) data sets included in submissions to regulatory agencies (e.g. FDA, EMEA). The SDTM data sets that become part of the submission are also the source for SAS programs used to generate analysis tables, listings, and figures that are the basis for claims of clinical safety and efficacy of the therapeutics being researched.

The CCI staff has two main roles: CDISC Consultants and CDISC Implementers.   Generally stated the Consultants advise on the mapping of clinical study data to the SDTM model, and the Implementers develop specifications and programs to implement the SDTM mapping.   A custom ETL tool named the Submission Data File (SDF) system is used to produce the SDTM data sets.

The CCI team are currently comprised of over 30 individuals working from Amgen sites and home offices located across the US, England, Columbia, and India. These teams now support over two dozen products and well over a hundred research projects (clinical studies). Each of these projects can have many milestones or deliverables: For example a project could have data review committee meetings, safety-update reports, interim analyses, primary analyses, and a final analysis occurring over a multiyear life-cycle. Each of these milestones requires the coordinated efforts of members of CCI staff who are often thousands of miles and many time-zones apart.

BUSINESS ANALYSIS

When evaluating different potential solutions to managing these projects and staff resources several factors and constraints were kept in mind. We were required to work with what was at hand: No big-budget tools were in the offing, nor would such be likely to be nimble or adaptable enough to meet all the team's evolving needs. Previous attempts to develop a solution using MS Project and MS Excel had proved unsuccessful. This was mainly because marshalling all the various project data was too big a task for any one project manager.   Multiple users updating shared common files proved unwieldy and fragile, and the so resistance to this approach waxed as confidence in the information it managed waned.   A new solution had to offer ease of information access and input so that team leaders could share the responsibility of making key project updates.   It had to be fairly good about letting multiple users make changes to the same data sets at the same time.   It had to be customizable, intuitive, and as

straightforward in getting data out as in putting data in. An audit trail, scalable security, and some means to recover data from 'accidents' were also key requirements.

Fortunately the teams were already using MS SharePoint for managing some common items such as meeting minutes, out-of-office calendars, and a variety of shared documents. MS SharePoint lists have the advantages of being as simple to implement as a spreadsheet while encapsulating powerful features such as role-based security, workflows, and web access. Since there was a level of familiarity with this product, it was available at no additional cost, and it had most of the desired features, the decision to implement our project plan data sets as MS SharePoint lists was a fairly easy one. For analyzing and reporting on this data the choice was likewise obvious: SAS is a core competency for our group and it was natural to seek to leverage this expertise. Some brief SAS code samples will be shared in an addendum to the paper to show how various example outputs were produced.

A key consideration was how SAS and SharePoint would exchange data. There are many options to accomplish this, from manual downloading to integration via SAS Business Intelligence Server.   After some false starts we developed a Java/Groovy utility that extracts from SharePoint and provides data files in a SAS readable format.   Starting with v9.3 SAS provides a 'GROOVY' proc that supports execution of Groovy code within SAS. Both this and the BI Server option offer additional benefits and would help automate the extraction process.   One of these options will likely be implemented in the near future as time and budgets allow.

The solution required a flexible and scalable data model. CCI resource forecasts are based on answers to three questions: What work do we do? What work needs to be done? Who will do the work?   The data model that was designed to support project resource supply and demand estimation is shown as a Venn diagram with three overlapping circles (see Figure 1), each circle being a data domain described further by child data sets in nested ellipses.   The overlapping areas of each domain represent shared information and the primary key fields that 'join' the data sets are called out on the connecting lines.   The next sections describe these data domains in more detail.
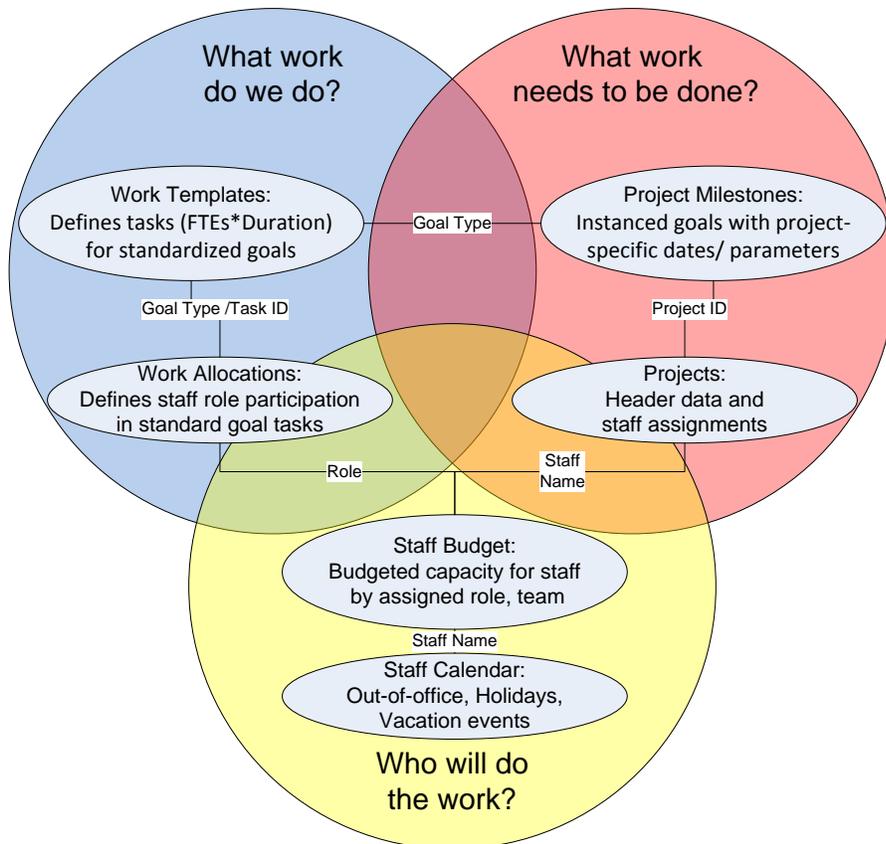

Figure 1: Resource Supply and Demand Data Model

DATA MODEL

Work can often be broken down into sets of standardized tasks that are needed for typical business objectives or goals, i.e. what drives the business, what it produces and how its success is measured.   These business goals are usually documented by whatever standardized operating procedures (SOP) govern the organization.   In modeling what work we do the SOPs and User Guides were our main reference materials.

The main deliverables and services provided by CCI are defined in by resource models termed 'goal types' in the Project Work Templates data set.   Each goal type is supported by a defined series of tasks with timeline and resource work load expectations and dependencies that are similar in concept to a project plans. In the following discussion therefore 'work plan', 'goal type' and 'template' are synonymous. Some examples of goal types for our team would be initial SDTM development, preparations of data snapshots for reporting, version migrations, etc. The advantages to abstracting the work plans in this way are that it is easy to refine these templates based on ongoing innovation and process enhancements while retaining the ability to apply such refinements across the current portfolio of work.

A separate data set 'Project Work Allocations' describes what roles take part in performing the standardized tasks in the work plans and to what extent. The CCI roles, Consultant and Implementer, share to some degree in the performance of most of these tasks. The Allocations data set therefore assigns a percent to each role for each task with the total allocation summing to and capped at 100%.

How is this approach useful? Consider a scenario that happened in our team recently: the group was reorganized so that some tasks were shifted to different roles. In this case the actual work tasks were not re-defined but the roles performing them were. It was a simple matter to change the Allocations table to assign different percentages to the roles to match this new structure.   When forecasting it is essential to be able to see not only how many staff will be needed but what kinds of staff (e.g. specialist roles, leads vs. regular staff) are required in order to balance loads effectively.   The data model provides this capability in a readily adaptable manner.

The 'Project Milestones' data set records specifics of instances of goal types for each supported project. Each project milestone is assigned a goal type, a start and/or a finish date, and a complexity factor as well as a number of other optional fields.   The complexity factor enables adjusting the duration and resource load for a milestone either up or down in a linear fashion.   The default complexity for any goal type is always '1'. What this 1 represents is that the default resource model is anchored at a duration / resource budget for a 'typical' activity of this type. What this means in practical terms is subject to change over time as practices evolve and efficiencies are gained through innovation and experience. If a given project milestone is rated at a complexity factor of 1.5 the assumption is that the tasks will be 1.5 times more time consuming and resource intensive than what is typical for that goal type.   For most goal types either the start or the finish dates are required. The resource estimation program can apply the template plan durations either forward from a start date or backward from a finish date depending on what was entered.   In addition to project milestones the 'Project' data set stores project header details including staff and study team representatives and a number of process support variables.

Staff capacity and the factoring of Out-of-office leave are the third component in the resource estimate equation. The budgeted time for each staff member is tracked in the 'Staff Resources' data set along with their role and main therapeutic area assignments. For a full-time employee the staff capacity is entered as 1 for each month they are expected to be actively employed.   For part-time workers or staff members that are on long-term leave the capacity is adjusted accordingly, e.g. a three day work week = 3/5 = .6 capacity.   This data set is only for recording budgeted capacity. Temporary leave due to sickness, vacation, or holidays is recorded in the monthly calendar shared on the CCI team home page. Each staff member is responsible to update this calendar with their Out-of-office events. Calendar events for holidays are populated by managers based on the approved corporate holiday list for the staff's country of employ.

All staff time-off events are calculated based on the individual staff members' budgeted capacity. This is done to ensure that leave time is not overstated, e.g. a part-time worker's leave does not reduce supply at the same rate as that of a full-time worker.   With this approach one can accurately plot the budgeted and available supply to identify planning challenges and get a reasonably accurate picture of near-term staff supply.

Why bother with all the overhead of holiday tracking? To illustrate how staff planning can be affected by local holiday patterns 2012 holidays for a sample of geographically distributed countries are shown in Figure 2. It should be apparent from this figure that remote workers in popular off-shoring locales may have cultural, national and seasonal vacation patterns that can cause unexpected impacts to staff supply planning.
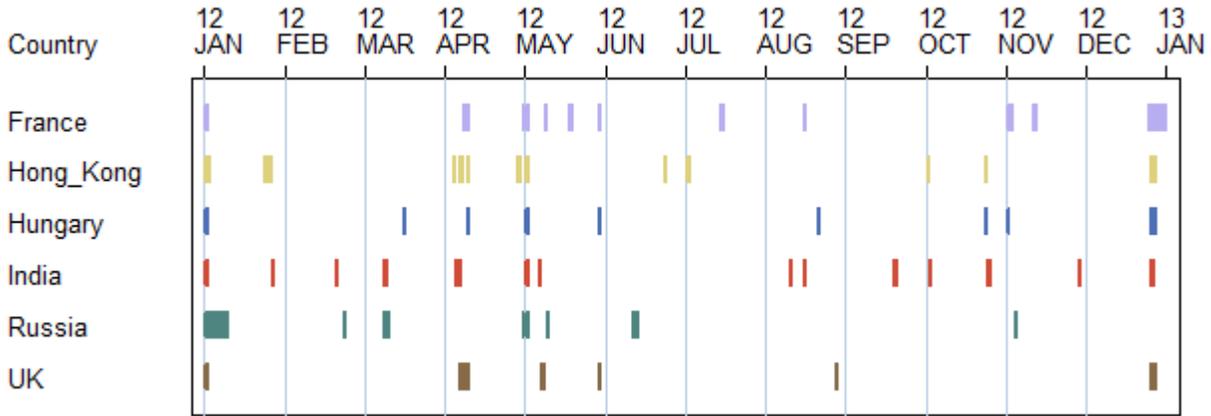
Figure 2: Comparison of 2012 Holidays in Six Countries

THE SOLUTION

So how do the data model pieces fit together to provide meaningful answers to the three basic questions: What work do we do? What work needs to be done? Who will do the work? It does so by enabling us to derive three key variables; demand, supply, and net supply. These variables are plotted across time in the sample graph in Figure 3.

Demand or work load is derived by programmatically applying the abstracted work plans to planned project deliverable or milestone instances. Some fairly complicated data step logic is needed to calculate the project work calendars in order to sum to the total expected full-time employees (FTE) needed on various dates. The total FTE demand appears as the green line. Budgeted supply, from staff resources, appears as the brown dashed line. Net supply equals the budgeted supply less staff holidays/leave and is shown by the dashed orange line. The histogram at the bottom of Figure 3 displays the 'gap' or variance between project demand and net supply numbers. This plot is a handy view into the overall resource needs for the group but doesn't really provide much insight into what product teams or projects are over or understaffed, or what roles with the teams may be disproportionately affected.
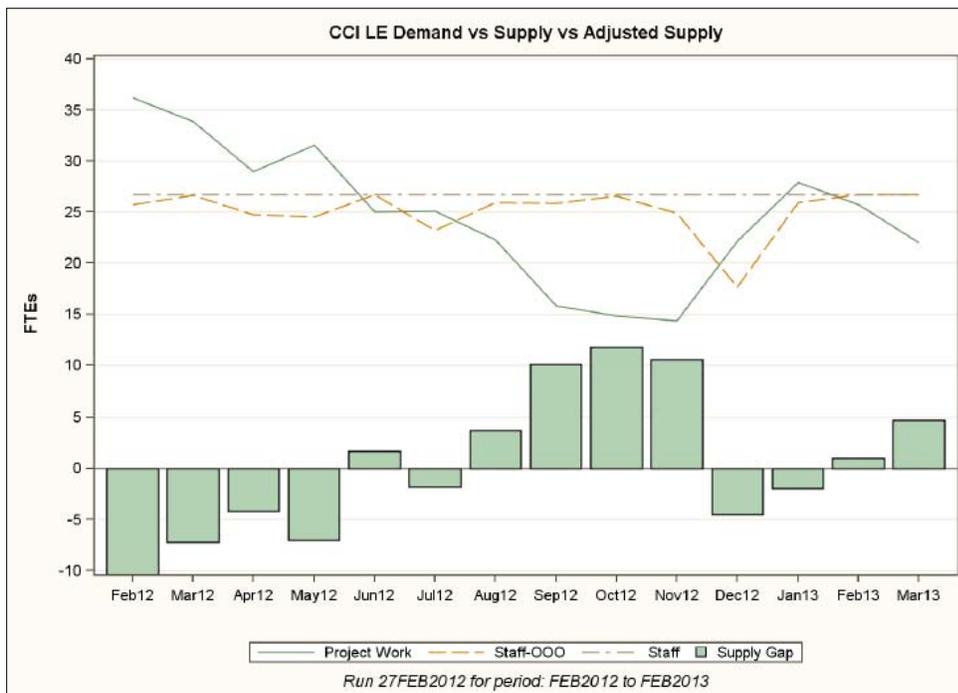

Figure 3: Sample Resource Projection Graph – Overall Demand vs. Supply

To gain insight on what teams and roles are impacted by planned work and to what degree a similar type of graph is rerun with grouping by Therapeutic Area (TA), a CCI sub-team, in Figure 4. In this chart the demand and adjusted supply and gap are broken out separately by role.
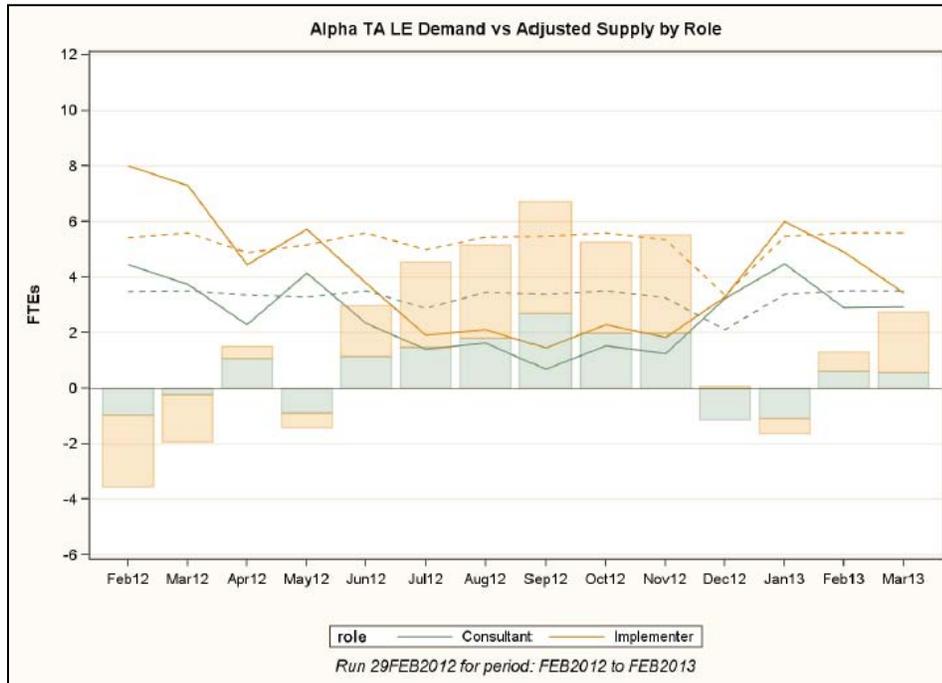

Figure 4: Sample Resource Projection Graph – Demand vs. Supply by Role for 'Alpha' TA

It is clear at a glance that some initial shortfall in staff in Q1-Q2 will become a surplus in Q3. In looking at the impact on roles in Team Alpha the need for Consultant support in May and December-January may be something that requires a shifting of resources.
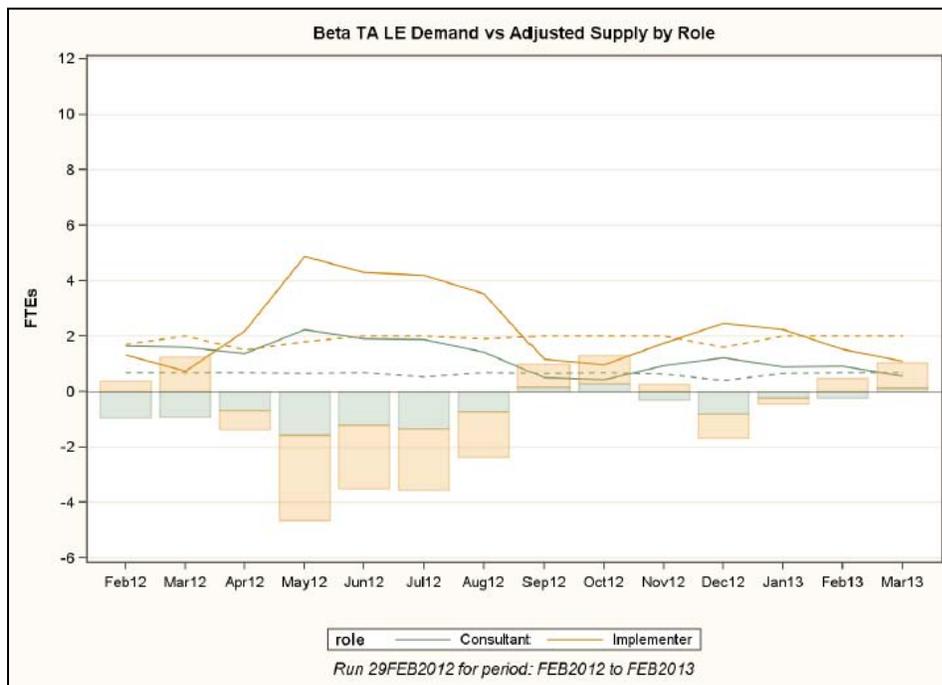

Figure 5: Sample Resource Projection Graph – Demand vs. Supply by Role for 'Beta' TA

The slowdown in work for the Alpha TA may be a blessing for the Beta TA as is apparent from a look at Figure 5. Here the increase in project work for the Beta TA creates significant staff resource gaps in Q2. Based on this a plan to cross-train staff from the Alpha TA to help with this bolus of work would be judicious. It is precisely these kinds of management decisions that prevent bottlenecks and/or under-utilization of staff and prove the value of investing in project management.

There are many other types of analyses and break-downs of the data that prove useful. For example the demand vs. supply can be looked at by product and by individual project or even by goal type deliverable. The adjustments to staff supply are looked at in detail in order to get early warning of potential gaps and the reasons why, e.g. a royal wedding or too many staff planning vacations at the same time. CCI decision making has also benefitted from the use of alternative outputs to MS Excel using the 'TAGSETS.EXCELXP' SAS ODS destination. The output workbooks provide the means to work with the data at the most granular levels while supporting filtering, sorting, and dynamic summation via formulas imbedded into the worksheets by use of PROC REPORT COMPUTE blocks and CALL DEFINE statements. An example of this output is shown in Figure 6.

	A	B	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Total Supply	Supply	29.1	29.7	29.7	29.7	29.7	30.2	30.2	30.2	30.2	30.2	30.2	30.2	30.2
2	Total Demand	Demand	-31.1	-25.9	-28.9	-31.3	-31.1	-26.3	-35.2	-33.4	-30.5	-33.4	-23.8	-26.0	-25.2
3	Net Supply	Gap	-2.0	3.8	0.8	-1.6	-1.4	3.9	-5.0	-3.2	-0.3	-3.2	6.4	4.2	5.0
4	-	Area Demand	AUG11	SEP11	OCT11	NOV11	DEC11	JAN12	FEB12	MAR12	APR12	MAY12	JUN12	JUL12	AUG12
5	Area	Alpha	-5.7	-4.6	-3.2	-4.6	-7.5	-5.4	-6.0	-8.2	-8.1	-8.0	-4.7	-9.4	-5.6
6	Area	Beta	-14.2	-10.6	-13.6	-19.0	-16.0	-10.1	-11.4	-9.2	-8.5	-9.0	-6.2	-4.0	-6.4
7	Area	Dweezil	-2.5	-1.5	-5.4	-2.9	-3.3	-3.4	-8.3	-4.1	-2.7	-4.7	-6.7	-5.5	-5.8
8	Area	Zappa	-8.8	-9.2	-6.7	-4.9	-4.3	-7.5	-9.5	-11.9	-11.1	-11.7	-6.3	-7.2	-7.4
9	-	Role Supply	AUG11	SEP11	OCT11	NOV11	DEC11	JAN12	FEB12	MAR12	APR12	MAY12	JUN12	JUL12	AUG12
10	Role	Conultant	8.5	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1
11	Role	Implementer	20.6	20.6	20.6	20.6	20.6	21.1	21.1	21.1	21.1	21.1	21.1	21.1	21.1
12	-	Product Gap	AUG11	SEP11	OCT11	NOV11	DEC11	JAN12	FEB12	MAR12	APR12	MAY12	JUN12	JUL12	AUG12
13	Product	Sugimab	0.3	0.0	0.3	-0.1	0.4	0.2	0.2	-0.5	-1.0	-1.2	0.0	0.1	0.1
14	Product	Globalex	-2.3	-0.5	-0.3	-0.1	0.0	-0.2	-1.0	0.4	0.5	0.2	0.5	0.5	0.2
15	Product	Forumica	0.0	0.1	-0.1	-0.2	-1.0	-0.2	-0.2	0.0	0.1	0.1	0.1	0.1	0.0
16	Product	Sasafrass	-0.2	-0.9	0.2	0.1	-1.3	-0.6	-0.6	-1.8	-0.4	0.1	-0.2	-2.1	0.0
17	Product	Pharmalot	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.3	-2.3	-1.9	-1.1

Figure 6: Sample Excel Worksheet Output of Resource Planning Data

A COMMUNICATIONS AND WORKFLOW HUB

The advantages of using this approach go well beyond centralizing project planning data. CCI uses MS SharePoint lists and workflows to organize many critical aspects of our business. One example is on-boarding and training records: This list is not only schedules and documents activities but also provides a library of useful links for students to their course materials. Many lists are used to aid group review and approval of metadata in the dozens of SDTM standard templates maintained in the SDF system. Important collaboration tools for issue tracking with project teams, managing development and changes to SDTM CRF annotations, tracking SDTM domain implementation, and conducting QC and locking tasks are also in use or in development. Some of the MS SharePoint lists are supported by sophisticated workflows that notify team members when their participation is required. We have also published a wiki site in SharePoint to provide guidance on our conventions in applying the SDTM model.

For our geographically and temporally diverse team this platform provides of ease of access, intelligent information sharing and maintenance, flexible asynchronous process support, and powerful reporting features. It has proven its worth and is now being extended to support new CCI staff and partners in a variety of locales.

CONCLUSION

SAS and MS SharePoint have been a winning combination so far but there are still many opportunities for improvement: CCI maintain data in over three dozen lists but currently regularly report out data from only six of these. Integrating this wealth of data into dashboards and other tools to help group decision-making is a near-term priority. As was mentioned PROC GROOVY and SAS Business Intelligence Server are two avenues we will be exploring to provide tighter integration and full access to this data. We have the normal challenges of evangelizing, educating, and supporting use of these tools in teams that have long histories with legacy processes. However the resistance to adoption has been minimal: Most teams quickly recognize the benefits over other communication and tracking strategies. By and large the biggest question has been not whether to use MS SharePoint and SAS but how to use them as these powerful tools provide so many workable strategies for meeting our needs.

REFERENCES

"The Submission Data File System - Automating the Creation of CDISC SDTM and ADaM Datasets", Marcus G Bloom and David Edwards, Amgen, Thousand Oaks, CA – PharmaSUG 2007.   Available at http://www.lexjansen.com/pharmasug/2007/ad/AD08.pdf

"Try This Demo: The ExcelXP Tagset and Microsoft Excel", SAS Institute Inc., Cary, NC. Available at http://support.sas.com/rnd/base/ods/odsmarkup/excelxp_demo.html

"Integrating SAS® Business Intelligence with Microsoft SharePoint", David R. Henderson and Sean Alexandre, SAS Institute Inc., Cary, NC – SAS Global Forum 2009. Available at http://support.sas.com/resources/papers/proceedings09/390-2009.pdf

"Study Data Tabulation Model Implementation Guide": Human Clinical Trials v3.1.2, CDISC Submission Data Standards Team, © 2008 Clinical Data Interchange Standards Consortium, Inc. Available at www.cdisc.org

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

W. Stetson Line
Amgen, Inc.
One Amgen Center Drive
Thousand Oaks, CA 91320
(805) 447-8326
sline@amgen.com
www.amgen.com

ADDENDUM - SAS CODE SAMPLES

For Figure 2: Comparison of 2012 Holidays in Six Countries

```
/* increment segment number for stacking holidays on one line per country and      _pattern to
vary colors by country */
data holidays;
 set holidays;
  by country date;
  retain segmt_no _pattern 0;
  if _n_=1 then _pattern = 10;
  start=date;
  finish=date;
  if first.country then do;
     segmt_no=0;
     _pattern = _pattern + 1;
  end;
  segmt_no=segmt_no+1;
run;

proc gantt data=holidays;
   id country;
   chart / ss=start sf=finish compress
          nolegend nojobnum
          ref='01JAN12'd to '31DEC12'd by month
           mindate='01JAN12'd
           maxdate='31DEC12'd;
   run;
```

For Figure 3: Sample Resource Projection Graph – Overall Demand vs. Supply

```
/* prepare data for SG Plot series. Define main y-axis supply and demand variables */
```

```
data sgseries;
merge zdemand(keep=area prod projid month role FTEs _way_ _type_ rename=(FTEs=demand))
      zstaff (keep=area prod projid month role FTES _way_ _type_ rename=(FTEs=staff))
      zooo   (keep=area prod projid month role FTEs _way_ _type_ rename=(FTEs=ooo));
  by area prod projid month role;
  attrib staff  label="Staff" format=8.;
  attrib supply label="Staff-OOO" format=8.;
  attrib demand label="Project Work";
  attrib supgap label="Supply Gap" format=8.;
  if ooo = . then ooo = 0;
  supply = staff-ooo;
  supgap = supply-demand;
run;


/* select overall totals from proc summary results, report data subset by global macro variables
defining reporting window */
ods proclabel  = "Demand Overall";
PROC SGPLOT DATA = sgseries description="Overall";
  where _way_= 1 and _type_ =2;
TITLE "CCI LE Demand vs Supply vs Adjusted Supply";
footnote "Run %sysfunc(putn(%sysfunc(today()),date9.)) for period: &rptdts to &rptdtf";
vline month / response=demand;
vline month / response=supply;
vline month / response=staff;
vbar  month / response=supgap;
yaxis values=(-10 to 40 by 5) grid label="FTEs";
xaxis label=" ";
RUN;
```

For Figures 4 and 5 – Demand vs. Supply by Therapeutics Areas grouped by Role

```
/* produce TA specific charts using by clause with area, within each chart group on role to split
out role specific supply and demand*/
ods proclabel  = "Demand: TA by Role";
PROC SGPLOT DATA = series_Ie description="";
by area;
TITLE "#BYVAL1 TA LE Demand vs Adjusted Supply by Role";
vline month / response=demand group=role lineattrs=(pattern=solid);
vline month / response=supply group=role lineattrs=(pattern=shortdash);
vbar  month / response=supgap group=role transparency=.7 name="Gap";
yaxis values=(-6 to 12 by 2) grid label="FTEs";
xaxis label=" ";
run;
```

For Figure 6: Sample Excel Worksheet Output of Resource Planning Data

```
/* assign MS Excel formulas round, sum, and sumifs to column styles*/
ods tagsets.excelxp file="&path.\reports\&rptdts._to_&rptdtf._SGP.xls" style=statistical;
run;

proc report data=sumrpt headline headskip spacing=0 missing nowd split="~" contents=""
style(header)=[background=verylightorange];
column id rowtype dummy2 dummy3 dummy4 label category role d0-d&nomos;
define id       / group noprint;
define rowtype  / display                        "";
define dummy2   / display                        "";
define dummy3   / display                        "";
define dummy4   / display                        "Summary";
define label    / display  style=[outputwidth=0.8in] "Class";
define category / display  style=[outputwidth=2.0in] "Category";
define role     / display                        "";

compute label;
    select;
    when('Total' =: label)
        call define(_row_,"style","style=[background=verylightgreen]");
    when('-'    =: label)
        call define(_row_,"style","style=[background=verylightgreenishblue]");
```

```
      when('Net'   =: label)
          call define(_row_,"style","style=[background=verylightvividorange]");
      otherwise;
      end;
endcomp;

%do i = 0 %to &nomos;
define d&i      / display format=4.1                    "";
%end;
%do i = 0 %to &nomos;
  %let c = %eval(&i + 8);
compute d&i;
  select(id);
  %do j = 1 %to &nsum;
    %let r = %eval((&j + 2)*-1);
    when(%eval(&j+0)) do;
    select(label);
      when('Supply')
      call define(_col_,"style",
          "style=[tagattr='formula:sumifs(R2C&c:R[&r]C&c,R2C6:R[&r]C6,RC6,R2C7:R[&r]C7,RC7)']");
      when('Area')
      call define(_col_,"style",
          "style=[tagattr='formula:sumifs(R2C&c:R[&r]C&c,R2C5:R[&r]C5,RC6,R2C7:R[&r]C7,RC7)']");
      when('Role')
      call define(_col_,"style",
          "style=[tagattr='formula:sumifs(R2C&c:R[&r]C&c,R2C7:R[&r]C7,RC6)']");
      when('Product')
      call define(_col_,"style",
          "style=[tagattr='formula:round(sumifs(R2C&c:R[&r]C&c,R2C2:R[&r]C2,RC6,R2C7:R[&r]C7,RC7),2)']");
      when('Total Supply')
      call define(_col_,"style",
          "style=[tagattr='formula:sumifs(R[&r]C&c:RC&c,R[&r]C5:RC5,RC6,R[&r]C7:RC7,RC7)']");
      when('Total Demand')
      call define(_col_,"style",
          "style=[tagattr='formula:sumifs(R2C&c:R[&r]C&c,R2C6:R[&r]C6,RC6,R2C7:R[&r]C7,RC7)']");
      when('Net Supply')
      call define(_col_,"style",
          "style=[tagattr='formula:sum(R[-2]C:R[-1]C)']");
      when('-')
      call define(_col_,"style","style=[tagattr='formula:R1C']");
      otherwise;
     end;
    end;
   end;
  %end;
  otherwise;
  end;
endcomp;
%end;
run;

ods tagsets.excelxp close; run;
```