ABSTRACT
Statisticians and programmers work in concert with each other to complete analysis projects and often resort to using Excel and email to store metadata, track project milestones, and collaborate with each other. See how using a single, centralized system to store TFL/Dataset metadata, communicate key analysis project milestones, report on QC efforts, conduct collaborative TFL reviews, and track TFL change requests can dramatically reduce data entry and reap big productivity benefits.

INTRODUCTION
Biometrics teams have many active analysis projects ongoing at the same time with multiple statisticians and programmers working on each one. Often, the work to store TFL and dataset metadata, communicate milestones, and TFL change requests happen in isolation within and across analysis projects.

Some teams have one or more in-house built tools to help automate tasks and communicate work, but most resort to Excel and email resulting in a lot of duplicate data entry and making it difficult, if not impossible, to report on.

In this paper we highlight the challenges statisticians and programmers face and share a novel solution to overcoming them.

Candid is an analysis lifecycle management system. It is a database backed web application that provides a single, centralized resource to store metadata, track analysis project milestones, and track TFL and dataset programming bugs and change requests. Candid is a hub for biometrics team communication and analysis project work.

WHAT IS THE MAIN UNIT OF WORK FOR BIOMETRICS TEAMS?
Our industry is rife with synonyms. Different words and acronyms (e.g. TFL, TLF, TLG) are used to represent the same thing (is it a complex algorithm, method, or comment?).

There are many common synonyms for analysis like deliverable, project, and study (study is often used interchangeably with analysis and clinical trial).

For the purpose of this paper we will use the term Analysis to refer to the main unit of work for Biometrics teams.

An Analysis is a discrete unit of work that has a plan (e.g. Statistical Analysis Plan), a team of statisticians and programmers, and a defined set of outputs (Tables, Figures, and Listings or TFLs) and datasets to be programmed, tested, reviewed and delivered to their endpoint.

TEAM USER SCENARIOUS WE WILL FOCUS ON
Using Candid’s workflow engine in concert with its metadata repository we will show how Candid provides a common resource for supporting key activities Biometrics teams do to complete analyses.
• Track analysis projects and communicate team responsibilities.
• Store, share, and manage TFL metadata.
• Store, share, and manage dataset metadata.
• Leverage TFL and dataset metadata to meet automation goals via an application programming interface (API).
• Track and report on TFL and dataset testing (aka QC or validation) processes.
• Securely share TFL files with statisticians and programmers.
• Conduct online TFL reviews.
• Track change requests, bugs, questions about TFL, dataset programming efforts.
• Get status updates at a glance from dashboards.
• Get metrics and reports on all above processes.

TRACK ANALYSIS PROJECTS AND COMMUNICATE TEAM RESPONSIBILITIES

OVERVIEW

User Goals

As a manager, I can’t easily see what my teams are working on without getting everyone to spend time writing status updates or looking at departmental or individual tracking spreadsheets.

As a programmer, I want to minimize my time updating my own and others’ tracking spreadsheets or sending e-mail status on my progress.

Solution

Managers, statisticians and programmers are all tracking their work, often separately in Excel. Shifting work from updating tracking spreadsheets to a central system means everyone knows where to go get their questions answered about project work. The nagging issues of locked Excel files, where spreadsheet trackers are located, and which one is correct instantly go away.

When teams are tracking their work in a centralized system rather than separately, everyone can see the results immediately. Figure 1 shows a section of Candid’s overview dashboard that shows how Candid automatically rolls up status details about active analyses, providing at-a-glance progress teams are making across the department.
Candid’s workflow engine helps productivity by automatically governing rules to move forward or back through the workflow and reducing the amount of data entry and formatting that spreadsheet tracking demands. All status changes are automatically date and time stamped and record the user who performed the changes. All these details are captured in an audit trail that tells the whole analysis story.

Analyses in Candid have a several workflows. This standardizes communication about project status, TFL programming progress, and dataset programming progress across the entire department. Figure 2 shows an example workflow for the conduct of an analysis.

**CANDID’S WORKFLOW ENGINE**

- Provides a consistent means of communicating Analysis, TFL, and Dataset progress.
- Provides a standard terminology for communicating progress.
- Enforces departmental workflow standards.
- Provides an automated audit trail (means less data entry over Excel or disparate systems to accomplish similar goals).
- Workflows can be customized per instance of Candid.

**ANALYSIS PROPERTIES**

In addition to project status, Candid also tracks the type, task, name, due dates, data sources, analysis team members (and their roles), and shared documents (e.g. Statistical Analysis Plan).
Table 1 Analysis Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Description, Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Ad-hoc, Planned</td>
</tr>
<tr>
<td>Task</td>
<td>NDA, Final, DMC Meeting, Publication, etc.</td>
</tr>
<tr>
<td>Name</td>
<td>A short, descriptive name for the project.</td>
</tr>
<tr>
<td>Data source</td>
<td>Associated clinical trial or trials.</td>
</tr>
<tr>
<td>Due date</td>
<td>Expected completion of delivering outputs.</td>
</tr>
</tbody>
</table>

THE ANALYSIS TEAM

Each analysis in Candid has team. Membership on a team serves two important purposes. The first is to communicate to the department the team members’ responsibilities. The second is to secure the analysis. A user’s combination of their membership and role on a given analysis determines what details they can see and what actions they can perform (e.g. a programmer can add or edit TFL metadata, whereas a reviewer cannot).

SHARING DOCUMENTS

It is easy to share analysis project related documents with the team. All team members (except Reviewers and Observers) can add and remove documents in the Supporting documents section on the detail page of the analysis.

STORE, SHARE, AND MANAGE TFL METADATA

OVERVIEW

User goals

As a statistician, I need to communicate TFL specifications and specification updates to programmers. I am eager to know when programmers have produced TFLs so I can begin reviewing them and providing my feedback as soon as possible.

As a programmer, I want to keep track of my programming progress and communicate my progress to statisticians and testers.

Solution

Storing TFL metadata and tracking progress in a single place allows has many benefits. Candid allows all team members to easily view and update metadata and progress. For example, a statistician may be monitoring the efficacy figures a programmer is working on. As soon as she sees the programmer has finished and uploaded them into the system, the statistician can immediately open the viewer and begin reviewing the TFLs. She can then quickly provide feedback back to the programmer, who will see that feedback on their dashboard (see Securely Share TFL Files with Statistician and Programming Team Members topic in this paper for more details).

All of this communication occurs inside the system. It is available to all team members, not hidden in email inboxes or spreadsheets.
Candid provides several ways to add, edit, and copy TFL specifications in bulk. There are wizards for copying existing metadata on to new analyses. Specifications can be exported to Excel, edited and then read back into the system. Candid also has an advanced templating system for storing standards that makes it very quick and easy to copy metadata in bulk. These standard libraries can generate titles and footnotes on new or existing analysis projects as well.

### STORE, SHARE, AND MANAGE DATASET METADATA

#### OVERVIEW

**User Goals**

*As a programmer, I want to keep track of my progress and see others’ progress on TFLs that are in progress.*

**Solution**

Centrally storing Dataset metadata and tracking progress through each Dataset’s lifecycle has many benefits. Similar to TFLs, teams track the programming effort and testing effort while at the same time storing SDTM and ADaM dataset specifications for each analysis. Storing variable details, controlled terminology (CDISC and custom), and value level metadata is also an option. Figure 4 shows an example of dataset details that include the workflow, top level dataset details and programming assignments.
As with TFLs, Candid provides several ways to input, reuse, and copy metadata to new or existing analysis projects. Dataset specifications can be exported to Excel, edited and then read back into the system. Candid also has an advanced templating system for storing standards that makes it very quick to copy metadata in bulk.

**LEVERAGE TFL AND DATASET METADATA TO MEET AUTOMATION GOALS VIA AN API**

Candid provides an extensive library of application programming interface (API) calls allowing programmers to access the metadata repository via an API. Whether you are using SAS, Python, R or any popular programming language, you have an interface to programmatically access the data and return either JSON or XML. You can also add and update data as well via the API. This opens the door for many automation opportunities.

Example using Candid’s REST API (for integration with other systems; call from SAS programs)

```sql
*-------------------------------------------------------------*
| HTTP to write XML file from Candid API                        *
*-------------------------------------------------------------*
proc http url="http://web_server/candid/api/v1/1/analysis"
  out=analysisList
  method="GET"
  ct="application/xml";
run;
```
SECURELY SHARE TFL FILES WITH STATISTICIAN AND PROGRAMMING TEAM MEMBERS

OVERVIEW

User Goals
As a statistician, I am waiting for the programmers to produce requested outputs so I can review them. This involves back and forth emails between stats and programming to figure out which outputs are ready and where they are located. Once reviewed, feedback is entered into spreadsheets or emailed to programming and is frequently not collated or easily tracked.

Solution
Once programmers have produced TFLs, the TFLs can be bulk uploaded into the system. This changes the status of the TFLs to Ready, alerting the statistician that the TFLs are available for viewing.

Statisticians (or other authorized team members) then use Candid’s Quick View utility to view TFLs and start providing feedback (e.g. change requests, questions, bugs) in the system by adding Issues.

Added issues show up on the programmer’s dashboard. The issues are automatically date and time stamped. Team members can see and respond to these issues directly in Candid.

Central tracking of Issues (change requests, questions, and bugs) provides the programming team a hub for discussing all the programming minutiae toward completing an analysis. Important findings, discussions about the best course of action, and the eventual resolution are preserved and searchable for future use.

CONDUCT ONLINE TFL REVIEWS

OVERVIEW

User Goals
As a statistician, I need to securely share TFLs with stakeholders in my organization, collect their questions and feedback, and then turn that feedback into action items for programming. I also need to monitor reviewer progress and make sure they are meeting the deadline. I typically have no visibility into the progress reviewers have made.

As a statistician, I have a big job at the end of a review to collate and make decisions about all the feedback. This feedback comes to me via tracking spreadsheets, email, and handwritten notes. I wish there was a system that would collate all those details for me.

As a reviewer, I want to know what TFLs I am supposed to be reviewing. I want to know how much time I have to provide my feedback, what details are important from the statistician’s perspective for me to focus on, and what has happened with the feedback I provided.

Solution
For most Biometric teams, conducting TFL reviews is a manual and laborious process. Candid’s review system makes it easy to conduct a formal TFL review. The entire process, from setting up a review to collecting reviewer feedback to reporting on actions taken with that feedback, is automated. Review periods and collected feedback are available to the whole team and are permanently archived. Figure 5 shows the workflow details for a single review period.
**With Candid’s Review Tool a Statistician can:**

- Identify which TFLs to share with reviewers.
- Select which stakeholders are allowed to view and comment on TFLs.
- Set and communicate a schedule for reviewers to provide their feedback.
- Monitor each reviewer’s progress.
- See and reply comments being added by reviewers in real time.
- Determine which comments are actionable or not.
- Provide instructions for programming that are automatically associated with the comment.
- Track programming progress toward completing actions items.

![Image](image.png)

**Figure 5. Review Period Workflow**

**With Candid’s Review Tool a Reviewer can:**

- See what review periods the reviewer has been granted access to.
- See the schedule for providing their feedback.
- View a summary of their own and others’ progress.
- See comments in real time and reply to others’ feedback.
- Add comments by marking up the TFL using drawing tools (e.g. pinpoint, a circle or square lasso, or highlight).
- See what happened to their feedback after the review period is over.

**Commenting**

Reviewers add comments in a variety of ways. Each comment is automatically date and time stamped and recorded with the name of the reviewer. Reviewers can also reply to comments from other reviewers to provide more information.

Comments can also be added by drawing on the surface of the uploaded TFL in Candid. Figure 6 shows the process to follow to add markup comments.
Figure 6. Reviewers can markup output inside the review tool

Increasing Reviewer Participation

Candid's review system has detailed dashboards that allow statisticians to monitor progress on the review period. Reviewer progress is automatically tracked by Candid. Figure 7 shows the reviewer progress chart on the review period's dashboard.

Figure 7. Statisticians Can Monitor Reviewer Progress

TRACK CHANGE REQUESTS, BUGS, QUESTIONS

OVERVIEW

Candid has a formal Issue tracking system that is available to all team members. An issue can be a change request, a question or a bug. Issues give a team a central, flexible resource for recording conversations about the specification and programming effort for analysis projects across the department. A single issue can be associated with both TFLs and Datasets.

An interaction on a single issue may look something like this:
1. Statistician creates a new issue for a programmer to address.
2. Programmer asks a question by adding a comment to the issue.
3. Statistician responds to question.
4. Programmer addresses the issue and records the issue as resolved and closed.
5. Statistician can see the issue is closed and reviews the newly created TFL to verify.

Centrally tracking issues means there are no spreadsheets to update and no buried emails to dig up.

**BENEFITS OF METRIC DASHBOARDS FOR TEAMS**

There are three dashboards in Candid. Each dashboard targets the primary questions different roles in Biometrics may have about departmental work.

**MAIN**

The Overview dashboard summarizes and shows high level metrics about all analysis work currently in the system.

**MY ANALYSES**

The My Analyses dashboard is the home page for statisticians and programmers. It displays a summary of each active analysis the logged in user is a team member of. Charts summarizing the workflow progress of TFLs and Datasets as well as links to open Issues are available here.

**REVIEW**

The Review dashboard is the homepage for all reviewers in the system. This page summarizes the logged in user’s access to current and past review periods they have participated in.

**REVIEW PERIOD DASHBOARD**

This dashboard serves as the homepage for every review period. Anyone on the analysis team can access review period dashboards for a given analysis. From here, team members can see a summary of comments and progress reviewers have made towards completing the requested TFL reviews. This is especially helpful for statisticians who (without Candid or a similar software) typically have little insight into how well reviewers are progressing through their review.

**GET METRICS AND REPORTS ON BIOMETRIC TEAM ACTIVITIES**

There are many predefined reports in Candid that summarize analysis projects. Reports can be viewed within the system and also can be exported to Excel. Table 2 Predefined Reports, lists a few of the reports that are available in all versions of Candid.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status Changes</td>
<td>Groups active analyses by their data source (Study).</td>
</tr>
<tr>
<td>Team Resourcing</td>
<td>Lists all users, their analysis membership and TFL and dataset assignment counts.</td>
</tr>
<tr>
<td>Analysis Lifecycle Report</td>
<td>Lists all outputs for a given analysis, their testing status and workflow histories.</td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>TFL Audit Report</td>
<td>Lists all TFLs, their programming, testing assignments and workflow histories.</td>
</tr>
<tr>
<td>Dataset Audit Report</td>
<td>Lists all datasets, their programming, testing assignments and workflow histories.</td>
</tr>
<tr>
<td>Issues Report</td>
<td>Lists all issues in a format that can be shared with team members or CROs outside of Candid.</td>
</tr>
</tbody>
</table>

We have been touting the benefits of avoiding spreadsheets. However, for those that need data from the Candid database, whether it be for internal use, trial master file input, or for sharing with external partners, exporting data to Excel is an option in many places with the click of a button. Analysis metadata, TFL metadata, dataset metadata, issues, user lists and user assignments are all examples of information that can easily be exported.

**CONCLUSION**

Candid provides the opportunity to have your team’s story all in one place. You can manage analyses, communicate team assignments, securely publish and review TFLs, manage all your output metadata, and automate tasks via the comprehensive API. As the pharmaceutical industry pushes to standardize, automate, and drive efficiency gains in drug development, Candid is a logical first step for biometrics departments that are still spreadsheet driven. Adoption of Candid can be phased in: first by replacing existing tracking spreadsheets then by driving deeper automation using centralized metadata, the workflow engine, and APIs.

We hope you found this paper interesting. Thank you for reading!

**CONTACT INFORMATION**

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

  Chris Hardwick, Justin Slattery, Hans Gutknecht
  Zeroarc, LLC
  info@zeroarc.com