PhUSE 2012

Paper RG03

Cost Effective Standards Implementation: A New Paradigm for the Drug Development Life Cycle

Jeffrey Abolafia, Rho Inc., Chapel Hill, NC USA
Frank DiIorio, CodeCrafters, Inc., Philadelphia, PA USA

ABSTRACT

Recent FDA guidances have made CDISC models the *de facto* standard for submissions. These standards are the foundation for systems that can improve workflow throughout the project life cycle. They create uniformity of metadata and data structures, which, in turn, provides an opportunity to write generalized, reusable code. When implemented properly, standards have tremendous potential for significant cost savings throughout the project life cycle. However, when implemented poorly, producing CDISC deliverables can actually *increase* the time and cost of drug development.

This paper outlines a new paradigm for cost-effective implementation of CDISC standards that is based on three principles: 1) Adopt a "Tables-first" approach: starting with the end instead of the beginning; 2) Extend standards to the real end, broadening the use of results level metadata and developing standardized display libraries; and 3) Integrate CDISC standards throughout drug development, using standardized metadata, libraries, programs, and tools.

INTRODUCTION

In recent years the FDA has clearly stated its preference for receiving both clinical and analysis data formatted in compliance with CDISC standards. This has been accomplished through a series of guidances, communication with sponsors, and presentations at conferences. As a result, CDISC models have become the *de facto* standard for submitting data to the FDA. In a draft guidance issued in February 2012[(1)], the FDA not only called for submitting standardized data, but also recommended that a data standards plan be submitted as part of the IND or IDE.

Given the FDA's preference for receiving CDISC data, many sponsors have begun to produce CDISC-compliant databases in order to meet FDA submission requirements. In the short term this has led to additional work and higher costs. However, when the standards are implemented properly, organizations have a tremendous opportunity for significant cost savings throughout product development. This paper lays out a framework for cost-effective implementation of CDISC standards that is based on three principles:

1) Adopt a "Tables-first" approach that starts with the end product (tables, figures, and listings) instead of the beginning (data collection);
2) Extend the use of standards further upstream to the protocol and downstream to the real end, broadening the use of results-level metadata and developing standardized display libraries;
3) Develop a data standards plan as part of the product development strategy.

THE MOVE TO CDISC

CDISC is an organization charged with developing standards that support the acquisition, exchange, and submission of clinical research data and metadata. It has developed standards that cover the entire life cycle of a clinical study, ranging from the study protocol to analysis and reporting.

The use of CDISC models has increased in the last few years. This trend shows no signs of slowing down. The PDUFA five-year plan includes the use of CDISC standards. The FDA is committed to using the CDISC SDTM and ADaM standards for data submitted to the FDA. The number of requests for CDISC submissions is increasing as many reviewers are developing a preference for and familiarity with CDISC standards and data. From a CRO perspective, there has been a marked increase in the number of requests for clinical data using the SDTM standard and analysis data modeled under the ADaM standard. In short, if you are in the clinical trial business, now is a good time to adopt CDISC standards and integrate them into your workflow.

Until recently, integrating CDISC standards and producing CDISC deliverables has primarily fallen on statisticians and statistical programmers. These two groups are largely responsible for mapping clinical data to SDTM, programming and validating SDTM datasets, creating define.xml file as documentation for the SDTM database, creating and validating ADaM datasets, producing the define.pdf file that documents the ADaM database, and populating the extensive amount of metadata required by CDISC standards.

Introducing CDISC models, especially SDTM, has a significant effect on work streams, work flow, and work processes [(3)]. Adding SDTM to the work flow means adding an entirely new work stream. The flow of work is no longer from the Data Management System (DMS) to analysis datasets. It is now from the DMS *to SDTM* to analysis datasets. This affects timelines, budget, and resources. In general, more of everything is needed.

TIMELINES

While there are now an increased workload and more deliverables to produce with SDTM, the metric for the number of days from database lock to top-line results and final displays has not changed. That being the case, more resources and improved coordination are necessary to meet project timelines.

NEW TYPES OF WORK

Producing SDTM and ADaM deliverables adds tasks to the pre-CDISC work flow. To create SDTM datasets and its required documentation:

- Specifications must be written to map data management data streams to SDTM domains
- Other SDTM required metadata at the dataset, variable, and value level must be populated
- SDTM datasets have be programmed and validated
- Trial design datasets have to be created and validated
- define.xml must be generated
- The CRF has to be re-annotated for SDTM
- SDTM datasets and define.xml must be validated to the SDTM standard.

Creating analysis datasets using the ADaM standard model also introduces new work streams, although not as far-reaching as with SDTM:

- ADaM datasets must be programmed and validated using SDTM as the data source
- The metadata requirements for ADaM add a significant amount of work to the pre-CDISC requirements. As with SDTM, metadata must be specified at the dataset, variable, and value level
- Results-level metadata, which documents displays and analyses, has to be populated
- ADaM datasets must be validated to the ADaM standard.

RESOURCES

This list of new deliverables and work flows underscore the down side of CDISC compliance. Incorporating SDTM, ADaM and other models adds new work streams and a significant amount of work, yet the timeline for producing analysis datasets and displays usually remains unchanged. As a result, more resources are needed to produce CDISC deliverables. Staff must be added and then trained to write SDTM specifications, design and create trial design datasets, populate the metadata required by SDTM and ADaM, program and validate SDTM datasets, re-annotate CRFs for SDTM, produce the required define files, and validate SDTM and ADaM deliverables.

LONG TERM BENEFITS

In the short term, integrating CDISC models into the work flow requires a sizeable investment. However, there are many potential long term benefits: 1) Using SDTM and ADaM as the corporate standard for clinical and analysis data provides an organization with standardized data across all studies and therapeutic areas; 2) Standardized data facilitates data exchange with multiple partners; 3) Standardized data facilitates data integration and producing the integrated summaries of safety and efficacy; 4) CDISC-compliant data facilitates review by regulatory agencies; and 5) Standardized data allows an organization to develop more efficient work processes and the potential for significant savings in time and cost.

POOR VERSUS SUCCESSFUL IMPLEMENTATION

The patent life of a new compound is now 20 years. Typically the time from patent to approval is greater than 12 years and *increasing*. That means that the time to recoup the investment in product development and make a profit is less than eight years and *decreasing*. Furthermore, the cost of developing new product has been estimated to be around one billion dollars.

So what does this have to do with CDISC implementation? If CDISC standards are implemented poorly in an organization, the result is likely to be more work, delays in timelines for single studies, delays in assembling a new drug application, a lower quality submission, delays in review and approval, increased costs, and ultimately less time remaining on patent.

On the other hand, a successful implementation of CDISC standards has the potential for: faster and more efficient study setup; shorter timelines; lowering the cost of product development; regulatory submissions that are easier to assemble and review; better communication during the review process; improved database warehousing and retrieval of information; and, most importantly, more time remaining on patent.

WHERE ARE WE NOW

The level of CDISC implementation varies widely across the pharmaceutical industry. However, it appears that the primary driver for implementing CDISC in the industry is to meet FDA requests for CDISC-compliant databases. SDTM and ADaM are the most frequently used models. CDASH usage is increasing but still not widespread, and the Protocol Representation Model (PRM) is relatively new and utilized by very few organizations. Thus it is safe to say that at this point CDISC standards are not part of an integrated product development strategy. Implementing CDISC standards remains primarily the responsibility of statisticians and programmers, and to a lesser extent, data managers.

Recent guidances and communication by the FDA have indicated that it plans to use CDISC standards for the foreseeable future. That means if you are in the drug development business you are also now in the CDISC and data standards business whether you want to be or not. Successfully integrating these standards into a product development strategy can not only get regulatory agencies what they want, but can also reduce the time and costs of drug development and approval.

HOW DO WE GET THERE

A successful, complete implementation of CDISC standards starts with a change of mindset in the problem we are trying to solve. The problem most organizations are trying to solve is "How can we get the FDA what they what?" Instead, we propose answering the question "How can we use CDISC standards as part of a cost-effective product development strategy?"

In this section we examine strategies for successful CDISC implementation. We propose a strategy based on three principles: 1) Adopting a "Tables-first" approach: start with the end instead of the beginning; 2) Extending the use of standards upstream to the protocol and downstream, to analysis and reporting; and 3) Creating a data standards implementation plan that is aligned with the production development strategy and integrates CDISC standards throughout the product development life cycle.

PRINCIPLE 1: START WITH THE END

The first principle calls for a tables-based approach where we start with the typical end point of product development, the displays for the ISS and ISE. Once we have a target product profile or a draft label, we design the mock displays needed to demonstrate safety and efficacy. Next, we determine what studies in a product development program are needed, based on the idea that each study contributes to the evidence required for approval. This method must be dynamic over the product development life cycle. That is, as we learn from early studies, we update the ISS/ISSE mock displays.

How does this work for a single study? First we generate the mock displays that are needed to meet the goals of the study. The mock displays serve as the driving force. They dictate the data to collect and identify the data streams needed. After the mock displays are created, we produce the protocol, the data collection system/CRFs, clinical data, the SDTM database, the analysis database, and finally the displays and statistical analysis.

Starting with the end may seem counterintuitive, but it has many benefits. Since we know all the studies we would like to carry out in advance, we can move on to the next study or phase more rapidly. (Later in the paper, we discuss how this is facilitated by implementing a data standards plan that covers the entire product development life cycle.) A second benefit is more focused data collection. We collect and validate only the data we need. Third, we can get an earlier start on analysis and displays. From the mock displays we know the analysis datasets required, the variables needed, and the analysis and displays that will be generated. Finally, this approach lets us get an earlier start on the ISS and ISE.

PRINCIPLE 2: END-TO-END STANDARDS IMPLEMENTATION

Most organizations have implemented standards with the goal of getting the FDA what they want. As a result, implementation has focused predominantly on SDTM and ADaM and to a lesser extent on CDASH. However, the use of CDISC standards can be extended both upstream to the Protocol and downstream to analysis and reporting.

Implementing at the Start. The CDISC Protocol Representation Model (PRM) identifies and defines over 100 elements that are common to clinical research protocols. The PRM standardizes many elements found in a typical protocol and renders them in machine-readable format. The PRM was developed so that information found in a protocol could be reused and repurposed across multiple documents and databases in a clinical study and across multiple studies and product development projects.

The PRM not only has the potential to streamline protocol development but can also facilitate creating databases and documents further downstream. The PRM enables setting up clinical data management systems (CDMS), allowing study design data to be exported from the protocol into the CDMS. Over 20 elements of the PRM map to elements in SDTM trial design datasets and to elements required for submitting to ClinicalTrials.gov. In addition the RPM can assist in preparing the following regulatory documents:

- PIND/PIDE meeting package
- INDs and CTAs
- Annual Reports/DSURs/PSURs
- EOP2 meeting package
- Pre-NDA/BLA meeting package
- NDA/BLA/Marketing Authorisation Application
- 120-Day Safety Updates

Implementing at the End. The use of CDISC standards can also be extended to the end of the project life cycle. Many of the data displays we produce are similar across studies and therapeutic areas. Many organizations are effectively using standardized libraries, programs, and metadata for the creation and validation of clinical and analysis databases. However, the use of these standardized inputs for analysis reporting lags behind that of database production. As the CDISC ADaM model has matured as an analysis dataset standard, the ADaM standard combined with standardized inputs can be utilized to facilitate producing displays and statistical analysis**.** At the 2012 FDA/PhUSE Annual Computation Science Symposium, the FDA identified this as a key initiative and put together a working group to develop standard scripts for analysis and displays.

By extending standards to protocol development and to producing statistical analysis and displays, we can realize the goal of end-to-end standards. In the next section we will examine the potential effect of end-to-end standards on product development.

PRINCIPLE 3: THE DATA STANDARDS IMPLEMENTATION PLAN

In a draft guidance issued in February 2012, the FDA recommended that a data standards plan be submitted as part of the IND or IDE. The plan should describe the standards a sponsor intends to use and how they will be implemented for each study. The guide also suggests that a data standard plan should be developed during the planning phase of product development, so that a sponsor can realize the full benefits of data standardization.

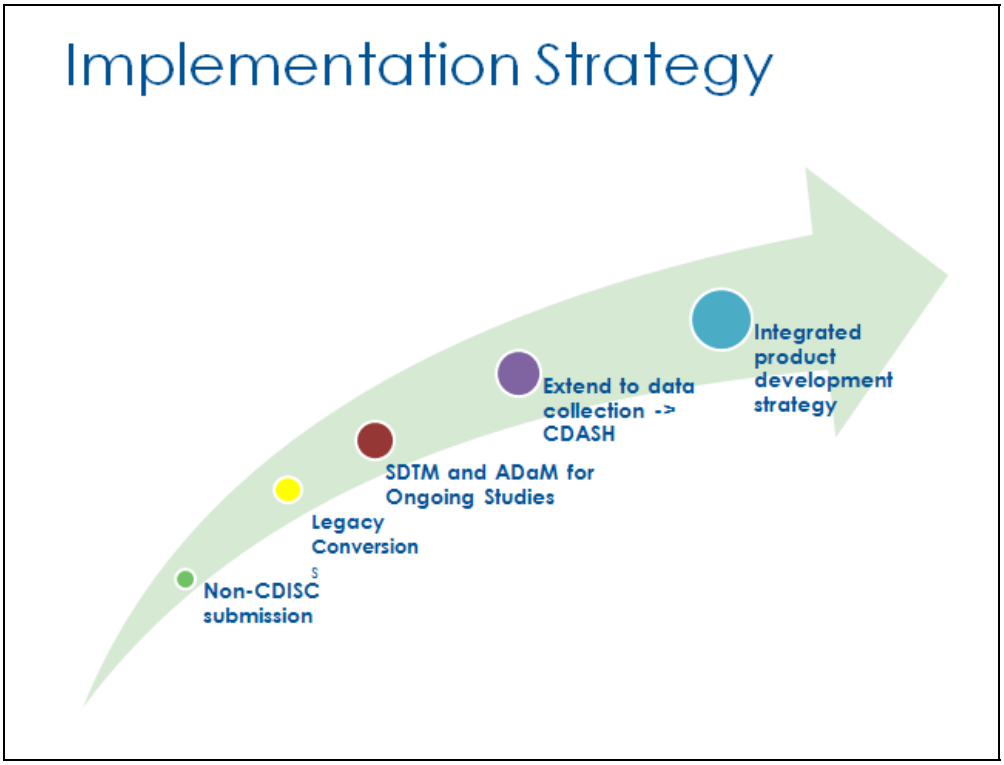
So what should a data standards implementation plan look like? To start with, one plan will not work for all sponsors or all products. A sponsor should consider the size and diversity of skill sets and what resources are available. The plan should also be in sync with a sponsor's business model. Is the goal to take the product to approval, to take the product to proof of concept, to partner as soon as possible, or to sell as soon as possible?

We now examine several implementation strategies and note the advantages and disadvantages of each strategy in regard to business model. These strategies are shown below in **Figure 1**, below, and range from no data standards implementation strategy to a fully integrated product development strategy.

NON-CDISC SUBMISSION

This strategy is listed because a CDISC submission is not yet an FDA requirement. Under this scenario, all datasets are submitted in a non-CDISC ("legacy") format and there is no attempt to integrate CDISC standards into an organization's work flow. We do not recommend this strategy because it does not get the FDA what they want. However, this is an option when you do not have the resources or the staff to produce CDISC-compliant deliverables.

FIGURE 1. EVOLUTION OF IMPLEMENTATION STRATEGIES



LEGACY CONVERSIONS FOR ALL STUDIES

This is where most organizations start. Under this scenario, the sponsor creates clinical and analysis databases using proprietary standards. At some point well after the study has been completed, typically after Phase Two or during Phase Three, the sponsor (or agent of the sponsor) converts the clinical and analysis databases to CDISC format. The clinical database is transformed to SDTM format and the analysis database is re-created under the ADaM standard, using the SDTM database as input. After the analysis database is re-created, a sample of the results is regenerated to ensure that the new ADaM compliant databases still supports the Clinical Study Report (CSR). These tasks are usually the responsibility of statisticians and programmers.

This approach is discouraged by recent FDA guidances but does have some benefits. It gives the FDA what they want: standardized CDISC-compliant data. This facilitates regulatory review and reduces FDA queries and the time to respond to them. A sponsor does not have to invest in CDISC until late in the product development life cycle. At this point the likelihood that the product will be successful is much higher than during earlier stages. Finally, a sponsor can continue to use proprietary standards as well as existing tools and processes.

This approach also has several severe drawbacks. It is expensive, and results in doing a good bit of work twice. The clinical and analysis databases, in addition to many of the displays, have to be recreated. Using this approach, a lot of work must be redone in a short period of time. This usually lowers the quality of the data and documentation, diverts resources from data integration and ISS and ISE, and in many cases delays the submission.

While not optimal, this is a good strategy if: the business goal is to sell as soon as possible or after proof of concept; the staff does not have the diversity of skills need to implement CDISC; or the sponsor does not have the resources necessary to implement CDISC early in development.

Legacy conversions are not a good strategy if the sponsor is working with multiple partners or the long term goal it to take the product to market. If a sponsor is partnering with several other organizations, there is likely to be a lack of uniformity across study databases. This makes data integration and data conversion to CDISC format even more cumbersome and resource intensive.

SDTM/ADAM IMPLEMENTATION FOR ONGOING STUDIES

Next in the progression is creating SDTM and ADaM datasets during the course of a study. Under this scenario: the operational database (data in the CDMS) is built using a sponsor's proprietary format; SDTM datasets are created from the operational data; the SDTM datasets are used to create ADaM compliant analysis datasets; and analysis and displays are generated from the ADaM datasets. As with legacy conversions, statisticians and programmers (and to a much lesser extent data managers) are still primarily responsible for standards implementation.

This approach not only gets the FDA what they want but is also a lot more efficient than the legacy approach. The clinical and analysis databases are only generated once; data for all studies are produced in a standard and uniform format, resulting in higher quality information; data is stored in an industry wide format; it is much easier to integrate data for the ISS/ISE, reducing the time and cost for preparing a submission; and the cost of producing analysis datasets and displays is reduced by utilizing standardized inputs.

This strategy presents several challenges in the short term. An organization must invest in changes to internal processes and workflow, training, and new tools and software. While an organization is getting up to speed on implementing CDISC standards there is more work to do, but not more time to do it in. This has the potential to delay study timelines and producing results needed for the CSR. Finally, most drugs (>90%) fail during Phase One, so there is the risk of doing a lot more work with little in return.

However, even though this approach requires a sizeable investment up front, we consider this is a step in the right direction if the business goal is to take the product to market or to partner. It does require a staff with a diversity of skills necessary to implement CDISC and adequate resources. Given that the FDA has invested heavily in CDISC in terms of tools and training and there is a proposed regulation to require standardized data, we feel that that investing in CDISC implementation is worthwhile in the long term.

SDTM/ADAM IMPLEMENTATION: THE HYBRID STRATEGY

This strategy is a combination of the two strategies we previously described. CDISC standards are applied only to adequate and well-controlled Phase Three studies, while data for all other studies remain in legacy format.

The advantage of this approach is there is a much higher probability at this stage that the drug will succeed. Also, there is a good chance you can negotiate with regulatory reviewers to submit only pivotal studies in CDISC format.

Among the shortcomings of this strategy: all studies will not be in a uniform format, which may annoy reviewers and increase the difficulty of review; integration will be more difficult (especially for safety data); and reviewers may prefer non-pivotal studies in CDISC format.

SDTM/ADAM IMPLEMENTATION PLUS CDASH

The strategies suggested up to this point are all aimed at producing the deliverables that the FDA wants. They are likely to result in additional work and increased costs. Additionally, they leave standards implementation predominately in the realms of statistics and programming. However, moving standards implementation further upstream has the potential to provide much greater value to an organization. The next two strategies we discuss extend standards to the start-up stages of a study.

The CDISC CDASH standard extends standards to clinical data management, with the goal of standardizing data collection. CDASH provides standard data streams and variables that are found in most clinical studies. CDASH was also designed to facilitate converting the operational database to SDTM.

CDASH provides a sponsor with a global library of data elements that are also the industry standard. The CDASH global library can be augmented by therapeutic specific libraries. These libraries can be utilized by the sponsor for all studies and product development project.

Extending standards to data collection provides many benefits. Using a global library of standardized data elements allows for cheaper and faster CDMS setup and conversion of operational data to SDTM. Standardized operational data combined with standardized programs, specifications and tools can streamline producing SDTM datasets. Producing the operational and SDTM databases can be packaged so that creating SDTM compliant clinical databases is cost effective for Phase One and Phase Two studies. This is a significant benefit for sponsors whose business goal is taking the product to market or partnering and for sponsors without a lot of resources in the earlier product development stages. Also, moving standards implementation upstream also increases communication among business units. Standards implementation is extended beyond programming and biostatistics to data management and clinical operations.

INTEGRATED PRODUCT DEVELOPMENT STRATEGY

The final and most comprehensive approach is developing a standards implementation plan that is integrated into the overall product development strategy. The plan will cover the entire product development life cycle, starting at protocol development and ending at the production of analysis and displays. Since a data standards plan is likely to be required at part of the IND or IDE, the plan might as well be part of the overall product development plan. The plan should outline which standards will be used, where in the product life cycle each standard is implemented, and how the various standards used will be integrated.

So, what would the standards implementation plan look like for a single study? We will build on the approach outlined in the previous section, extending standard implementation to the protocol phase. We will start with the mock displays for a given study. Most of our mock displays will be derived from a global library of mock displays. This library will contain displays typically used for all therapeutic areas, therapeutic specific displays that are used frequently, generic programs that generate the displays, and metadata or specifications that document each display.

Next we will generate the protocol using the Protocol Representation Model (PRM). This process will be streamlined by utilizing a library of common protocol elements that can be repurposed across studies and products. The study design and study schedule will be exported from the PRM to facilitate setting up the CDMS. We will then setup the CDMS using a global CDASH library that consists of forms, data elements, and validated edit checks.

Using CDASH compliant operational data, our library of standardized programs and specifications, and output from the PRM, the operational data will be converted to SDTM format. The SDTM database (along with a library standardized programs and specifications) will be used as the source for ADaM compliant analysis datasets. Finally, a library of standardized programs will be utilized to produce the analysis and displays.

At the completion of a study, we will have CDISC compliant clinical and analysis databases that are submission ready. These databases will facilitate integration of data from other studies because they are standardized. We will also be able to draw on data in the PRM to enable preparing other regulatory documents, submitting to ClinicalTrials.gov, and generating the protocol for future studies.

It's hard to imagine discussing anything related to submissions without including regulatory personnel. This approach finally broadens standards implementation to regulatory and clinical staff. This promotes cross-functional communication and planning at the early stages of product development. It also enables data integration, assembling the ISS/ISE, producing other regulatory documents, and regulatory review.

End-to-end standards implementation is most cost effective when the goal is to bring the product to market. This strategy also adds significant value when the goal is to partner or to outsource to multiple vendors. Finally, standardized data is more valuable to a merger or acquisition partner.

A business case study on CDISC standards by Gartner[(2)] found that implementing standards from the beginning can save up to 60% of non-subject participation time and cost and that about half of the value was gained in the startup stages. The study also reported that the average study startup time can be reduced from around five months to three months.

CONCLUSION

It is likely that a data standards plan will soon be required by the FDA at IND or IDE time. That being the case, we recommend developing a data standards implementation plan as part of a product development strategy. The plan will be dependent on business strategy, company size, and resources available. A successful implementation strategy can reduce the time and cost of product development, facilitate regulatory review, increase communication, increase data quality, improve internal processes, and increase the time remaining on patent.

Developing a successful strategy requires a change in organizational philosophy. We propose starting with the end to expedite product development. We also outline an end-to-end approach that extends standards beyond programming and biostatistics to data management, clinical, and regulatory staffs, utilizes standards from protocol and displays, and makes standards implementation an integral component of product development.

If you are in the pharmaceutical product development business, you are now also in the data standards business!

REFERENCES

(1) U.S. Department of Health and Human Services - Food and Drug Administration (2012): Guidance for Industry – Providing Regulatory Submissions in Electronic Format – Standardized Study Format: Draft Guidance.

(2) Rozwell *et al.*, 2009, online at http://www.cdisc.org/business-case. A business case for standards by CDISC and Gartner.

(3) Abolafia, Jeff and Frank DiIorio, “Brave New World: How to Adapt to the CDISC Statistical Computing Environment”. 2011. Proceedings of the Pharmaceutical SAS Users Group Conference.

ACKNOWLEDGMENTS

We would like to thank our colleague at Rho, David Shoemaker, for his valuable input during preparation of this paper. As always, we would also like to acknowledge the proofreading expertise of April Sansom.

CONTACT INFORMATION

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

Jeffrey Abolafia
Rho, Inc.
Jeff_abolafia@rhoworld.com

Frank DiIorio
CodeCrafters, Inc.
Frank@CodeCraftersInc.com