PhUSE 2014

Paper CD10

Every Study is Special! - Governing Data Standards

Chris Price, Roche, Welwyn Garden City, UK
Elizabeth Nicol, Genentech, South San Francisco, USA

ABSTRACT

The definition of data standards aligned with CDISC is just the start of a journey. Once standards have been defined, they need to be implemented on each study according to the protocol in data collection, tabulation and analysis. Roche governs the Global Data Standards for all of these areas with a cross-functional committee managed by the Data Standards Office (DSO). The committee is made up of five core teams each supporting a different area of the standards. Each study team, where necessary, submits requests for approval for deviations from the current standards to be implemented at a study level. In addition to deviations for individual studies requests are also received for the development of new standards. All requests are evaluated by the DSO and then brought before the Governance Committee for a decision and actioned as appropriate.

INTRODUCTION

The usefulness of having an internally aligned set of defined data standards from data collection to data analysis is well understood. Standards allow for a consistent approach to collection, tabulation and analysis for these activities and assessments which are routinely performed as part of a clinical study, allowing time to be spent on more innovative and value-added aspects of clinical studies. With the advent of CDISC standards, especially as they have matured over recent years, it is an obvious progression to align internal data standards with the published Clinical Data Interchange Standards Consortium (CDISC) standards. The recent publication of the draft FDA "binding" guidance relating to the submission of standardized study data has moved this from a beneficial "nice to have" to a business critical requirement.

Developing and defining a set of Global Data Standards is the start of a much longer process. Following the initial definition of a set of standards, these need to be implemented, their usage enforced, and then maintained so that they keep pace with advances in clinical knowledge and clinical practice. How this is accomplished is extremely important. There is a need to balance the efficiencies gained from data standards with the need to address new scientific questions. While significant elements of a study can be re-used from one study to another or selected from a subset of a master set of standards, each study has its own unique elements that need to be supported. Applying governance to these elements allows reuse in subsequent studies and ensures that all downstream impacts are appropriately considered. A key part of governance is to ensure that the most appropriate people are part of the decision making process so that a decision can be supported and subsequently enforced.

HISTORY OF STANDARDS DEVELOPMENT AT ROCHE

The use of standards has a long history at Roche and Genentech. Prior to the merger of Roche and Genentech, each company managed data collection and data tabulation standards as separate entities according to each company's respective internal models. Following the merger in 2009, the DSO was created in the combined Biometrics organization to lead the Global Data Standards Integration Initiative (GDSII) to develop a universal set of data standards aligned where possible with CDISC, across all therapeutic areas within the existing pipeline.

The GDSII consisted of three different streams addressing data collection, data tabulation and data analysis. The Data Collection and Data Tabulation streams worked closely in parallel towards the development of standard Case Report Forms (CRFs) pages and their associated Study Data Tabulation Model (SDTM) annotations, consulting widely with a cross functional team of subject matter experts. The Data Analysis stream also worked with a team of cross-functional subject matter experts in the development of standard algorithms to be implemented in all Statistical Analysis Plans. An important decision taken during the GDSII was to develop a metadata registry (MDR) that is underpinned by semantic technology. This MDR is now in production and allows the DSO to efficiently maintain the Global Data Standards and make them readily available to users and applications.

Since the completion of the GDSII, significant efforts have been invested in the maintenance of the existing Global Data Standards and in the development of new standards to support the ongoing needs of the business. In addition to these day-to-day tasks, the DSO has become involved in more strategic projects helping to develop tools driving the automation of certain study activities, taking advantage of the use of Global Data Standards.

THE DSO

Since its initial creation in 2009, the DSO has increased in both number, and the type of roles. A team of twenty, from a variety of backgrounds, is currently supporting the development, maintenance and adoption of Global Data Standards at Roche. There are three distinct teams each having a different specialism: Data Collection; Data Tabulation and Analysis; and Information Architects. Although data collection, tabulation, and analysis are supported by different teams of Global Data Standards Managers, coordination between these is vital to ensure all Global Data Standards are approached from an end-to-end perspective. Updates in one area inevitably have an impact elsewhere. The Information Architecture team is responsible for the maintenance of the Global Data Standards Repository (GDSR), the Roche sematic technology based metadata registry. Currently there are thirteen Global Data Standards Managers and six Information Architects working in the DSO plus a Global Head.

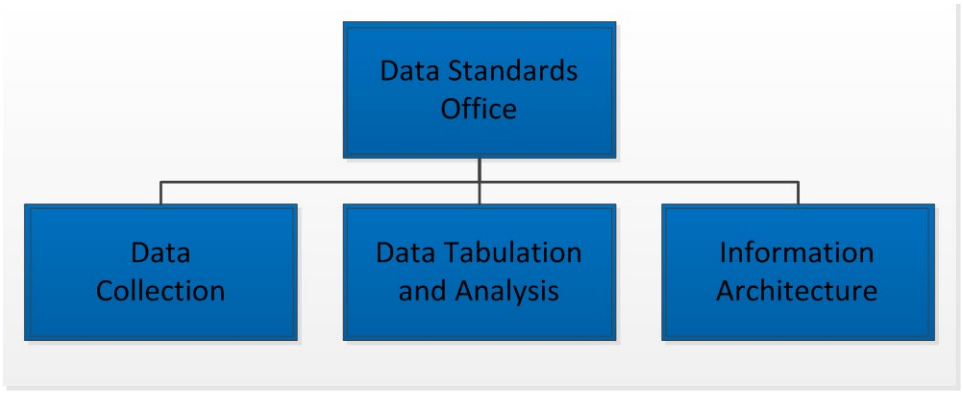

Figure 1: DSO Structure

GLOBAL DATA STANDARDS

Global Data Standards have been developed at Roche across data collection, data tabulation, data analysis and controlled terminology. These were developed by cross-functional teams of subject matter experts and have been continually maintained since their initial development. The Global Data Standards are made up of an information layer and in some cases an implementation layer. The information layer defines the "What" in relation to the Global Data Standards, independent of any system constraints. The implementation layer describes the "How" and is made up of tool dependent metadata describing how the information layer is implemented in practice. The separation of these two elements allows for tools and systems to be updated with no impact on the data standards with only the "bindings" between the standards and the system/tool requiring updates.

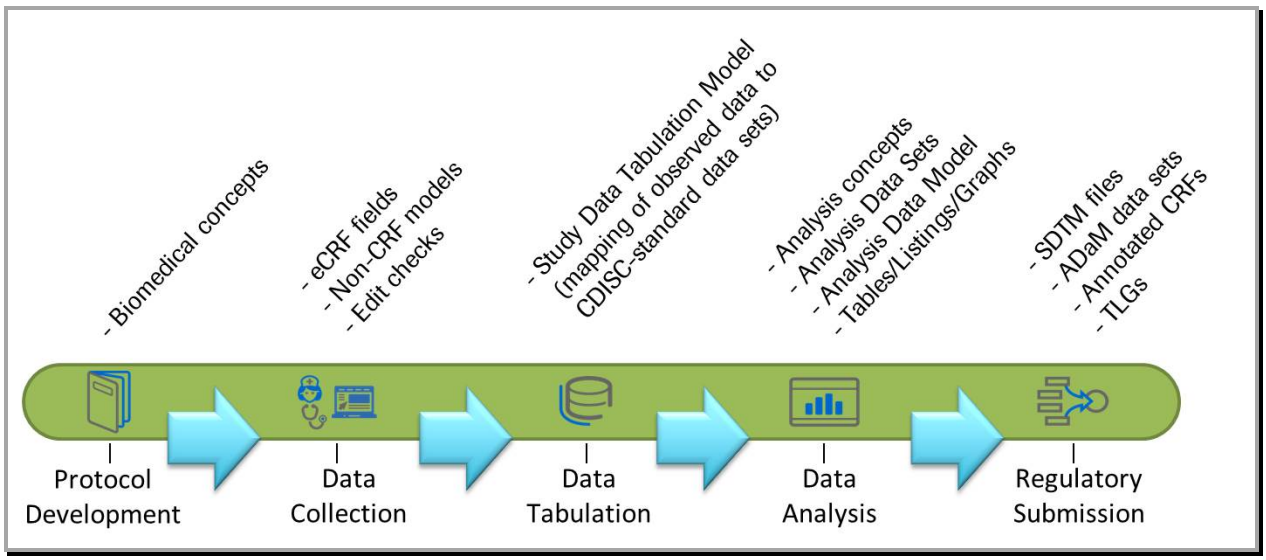

Figure 2: Global Data Standards

DATA COLLECTION

The data collection standards comprise of both eCRF and non-CRF data elements that are used to collect clinical trial data. The information layer of the eCRF and non-CRF data collection standards is constituted of biomedical concepts and data elements that are organized by activity. In this instance an activity is defined, in line with the NCI

thesaurus, as an action, undertaking, or event, which is anticipated to be performed or was performed according to the study protocol during the execution of the study.

The eCRF data collection standards also have an implementation layer component. They include a set of system-dependent implementation rules (e.g. data modeling rules, edit checks) that are used to build the Global Data Standards as part of an Electronic Data Capture (EDC) global library volume. The EDC global library volume is a secure and fully tested reference library that is used internally to copy standard elements such as fields, forms, and edit checks to enable the build of individual clinical study databases.

DATA TABULATION

The data tabulation standards are constituted of two related components. The first component is the data element level annotations of all the data collection standards describing where each is mapped to within SDTM. The second component is the data tabulation model based completely on SDTM. This describes the structure of the data tabulation datasets. Further to the metadata contained within the SDTM Implementation Guide and defined supplemental qualifier variables, additional content has been defined relating to variable lengths, a core status from a Roche perspective and implementation notes providing additional Roche specific information to the CDISC notes.

DATA ANALYSIS

Data analysis standards include both an information and an implementation layer. The information layer is made up of analysis concepts and shells for tables, listings and graphs. Each analysis concept is an algorithm describing how to derive an analysis endpoint from the simple (e.g. study day) to the complex (e.g. date of progression/censoring). These describe the algorithm in non-technical language so that they can be understood and reviewed by those outside of Biometrics. The table, listing and graph shells include both statistical concepts (e.g. mean, standard deviation, median) and analysis concepts to be included in a particular output and the structure of the output.

The implementation layer includes more technical information specifying how the analysis concepts and table, listing and graph shells should be created. These include analysis dataset specifications aligned with the ADaM Implementation Guide plus Roche extensions relating to additional variables as well as variable attributes not specified in the implementation guide. Additional Roche extensions include data models for analyses that are not supported by ADSL or the Basic Data Structure (BDS). Each derived variable and/or parameter within the standard analysis datasets must have a supporting analysis concept associated with it. Technical specifications for the tables, listings and graphs are dependent on the particular shell but regularly include information relating to the sort order, programming options in the relevant programming language (usually SAS®) and criteria for inclusion of data in a particular analysis.

CONTROLLED TERMINOLOGY

Controlled Terminology is applicable across all areas of the Global Data Standards from data collection to data analysis. By design CDISC Controlled Terminology is included as part of the data collection standards and is not updated within the data tabulation standards, with the exception of where synonyms have been used in data collection to simplify the provision of data (e.g. Complete Response in data collection is a synonym of CR in data tabulation). Similarly, no updates are made to the tabulated controlled terminology within data analysis, although at each stage additional controlled terminology is available for derived or assigned values. Controlled terminology is managed separately within the Global Data Standards with input provided by each of the teams. All implementations of controlled terminology are managed by each group accessing the same master list of controlled terms.

GLOBAL INFORMATION STANDARDS GOVERNANCE COMMITTEE

All Global Data Standards and how they are implemented are governed, by a cross-functional Global Information Standards Governance Committee (GIS-GC). This committee is made up of four levels: An Advisory Board, Across Therapeutic Area Clinical Core Teams, Therapeutic Area Standards Experts and an Extended Team made up of ad-hoc subject matter experts. In addition to governing Global Data Standards further core teams exist to assist in the governance and usage of “Master data” across the Product Development (PD) organization, details of which fall outside of the scope of this paper.

The committee’s high-level objective is to standardize clinical data and information, from its inception in clinical development plans, study designs, protocols and data analysis plans to its collection, tabulation, analysis and submission. This drives greater semantic interoperability between colleagues, processes and systems to simplify the current state. The committee supports the development of new Global Data Standards where none existed previously, the revision of existing Global Data Standards and the implementation of Global Data Standards by project and study teams.

Members appointed to the committee are expected to represent their own functions. If they believe a colleague can offer a greater level of expertise on an item the committee is addressing, they pro-actively reach out for input so that it can be shared with other committee members. In addition to their role on the GIS-GC, members also can represent Roche on external initiatives (e.g. Transcelerate BioPharma, CDISC, and PhUSE CSS Working Groups) as well as providing feedback on proposals from Standards Development Organizations (SDOs).

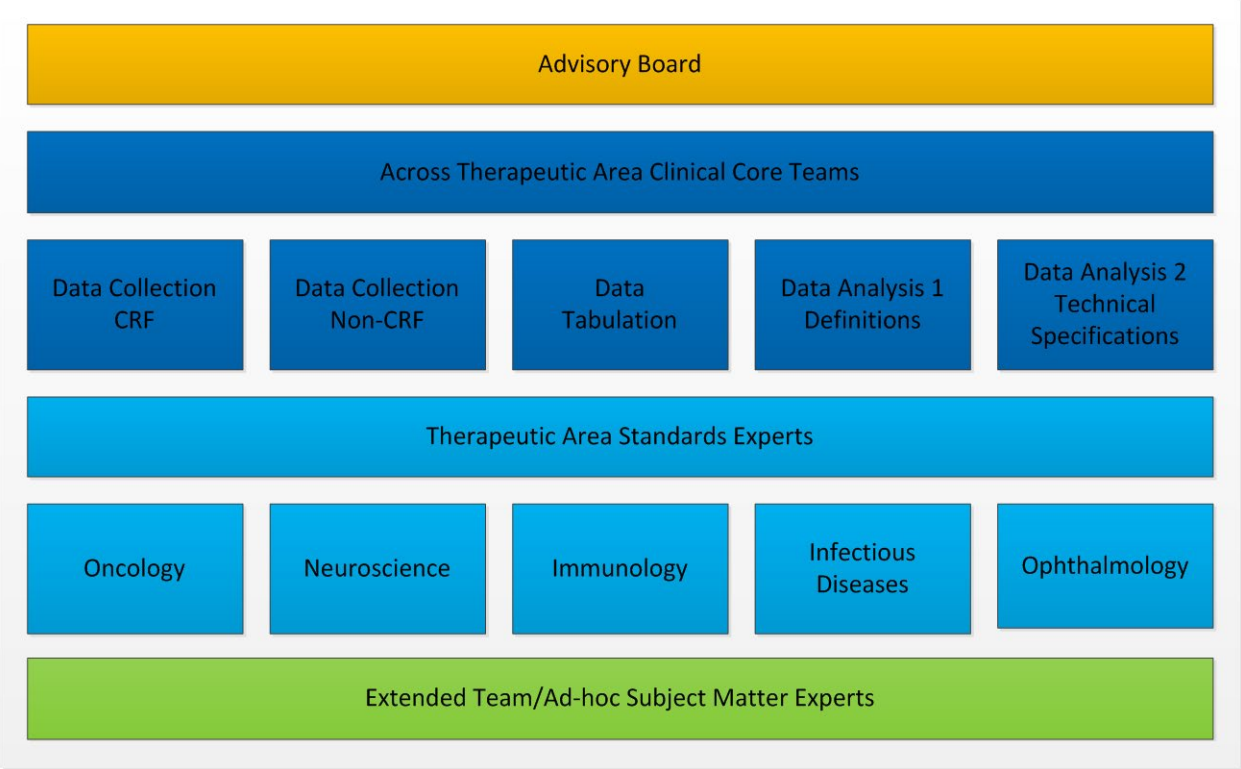

Figure 3: Global Information Standards Governance Committee Structure

ADVISORY BOARD

The Advisory Board for the GIS-GC is comprised of senior leaders from across the PD organization and its partners, including Diagnostics and Informatics, who have accountability for data and information standards within their given function. The Advisory Board provides strategic direction for standards development and helps to prioritize standards development activities. It also acts as the final decision making body for core team leads to escalate cross-functional disputes that may arise during standards development activities. The Advisory Board meets quarterly, or as required to discuss escalated items.

ACROSS THERAPEUTIC AREA CLINICAL CORE TEAMS

There are five Across Therapeutic Area Clinical Core Teams, each supporting a specific area of standards development. There are two teams supporting data collection (CRF and Non-CRF), one supporting data tabulation and two supporting data analysis (Definitions and Technical Specifications). Each core team has a clear area of responsibility and consists of cross-functional subject matter experts in that area. Members of the DSO lead each core team. They serve as a bridge linking the core teams together to support the end-to-end development of data standards. The core team lead is responsible for coordinating discussions, gathering input and acting as the final decision-maker for all issues raised to the particular core team. The core team lead does have the option of escalating issues to the Advisory Board in the (very rare) scenario of a strong disagreement within the core team, where solid contradictory rationales have been provided by the functional subject matter experts. The Across Therapeutic Area Clinical Core Teams meet weekly to allow for the timely resolution of all Global Data Standards related matters.

THERAPEUTIC AREA STANDARDS EXPERTS

The Therapeutic Standards Experts support the core teams in the development, maintenance and implementation of therapeutic area and indication specific standards; this includes both the information and implementation layer of the Global Data Standards. In particular they are responsible for the continued development of new Global Data Standards and the review of existing Global Data Standards within their therapeutic area or indication of expertise, to ensure they are appropriate for use by study teams.

EXTENDED TEAM/AD-HOC SUBJECT MATTER EXPERTS

The Extended Team are ad-hoc subject matter experts spanning every therapeutic area and key functions. They support the Core Teams by providing their expertise on an as-needed basis to ensure that all functions and therapeutic areas are fully represented in the decision making process. They also act as subject matter experts for specialist areas that require a continuous flow of standards development activity (e.g. laboratory data, biomarkers, PRO data).

DEVELOPMENT AND MAINTENANCE OF GLOBAL DATA STANDARDS

A key role of the GIS-GC is to develop and maintain the Global Data Standards. This may be driven by many different factors; these may include, but are not limited to:

- A compound enters clinical trials for a therapeutic area or indication not previously addressed by the Global Data Standards.
- Scientific or statistical understanding increases or changes leading to new or enhanced endpoints for clinical studies and the Global Data Standards need to be updated.
- A pattern is detected from study-level Global Data Standards requests identifying existing Global Data Standards that may require updating.
- A new data model (or new version of an existing data model) is developed and needs to be adopted as part of the Global Data Standards.

The entire GIS-GC supports the development and maintenance process with each constituent part providing input and expertise as required.

DATA COLLECTION

For the development and maintenance of Global Data Collection Standards a clearly defined process has been developed. Principally this is applicable to therapeutic area standards but can be applied to the more stable across-therapeutic area standards. The process is split into distinct phases and is led by the Therapeutic Area Standards Experts with each member providing their specific area of expertise at the relevant stage of the process.

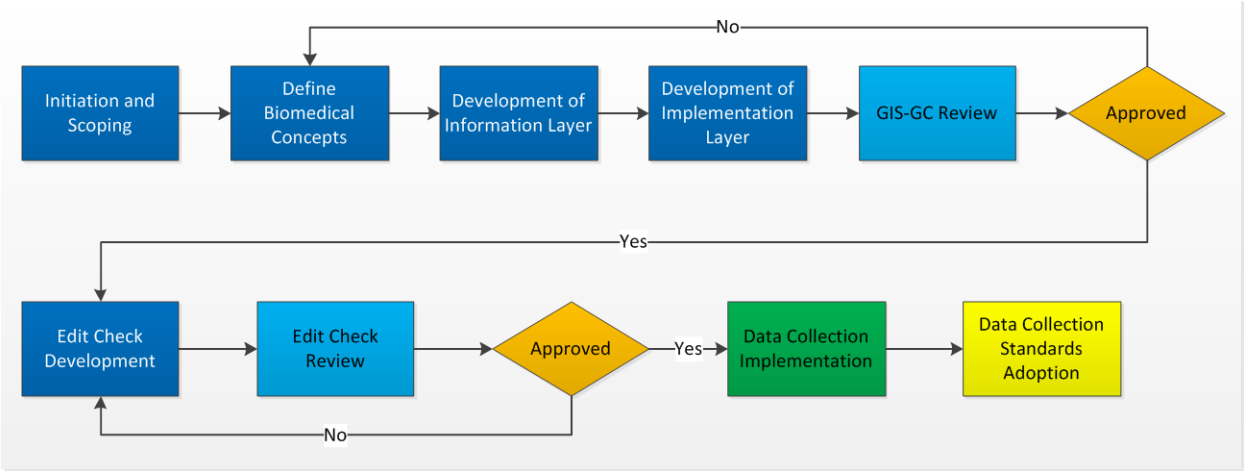

Figure 4: Global Data Collection Standards Process Flow

The initiation and scoping, definition of biomedical concepts and the development of the information layer are mainly supported by Clinical Science and Biostatistics representatives. The initiation and scoping phase consists of the team kick-off and the decision as what study activities and modules are to be included in a particular round of standards development. This decision will be driven by the priorities of the business and will not necessarily lead to a complete set of standards for an entire therapeutic area or indication. It is at the initiation and scoping phase that research will be done to establish what standards already exist for the therapeutic area or indication externally (e.g. CDASH, CFAST TA User Guides). Where standards already exist internally these should be included within the scope and evaluated to ensure that they are still meeting the requirements of the therapeutic area or indication.

The definition of biomedical concepts and development of the information layer phases both involve additional details and granularity to the high-level scope defined during the initiation and scoping phase. At both stages every effort is made to reuse existing Global Data Standards that may be applicable from other therapeutic areas or indications. Data elements are defined as part of the development of the information layer stage; these are individually collected data points that may be required for analysis within the therapeutic area or indication. Also included as a property of a data element is the type of response expected (e.g. numeric, controlled terminology or free text). In the case of controlled terminology the CDISC or Roche code list should be assigned. If a CDISC code list does not exist then the DSO representative will submit the code list and/or new values to CDISC for possible addition to the CDISC controlled terminology standards. Biomedical concepts are logical groupings of data elements; these groupings are normally formed based on medical rationale. For example, a biomedical concept of blood pressure might contain the data elements diastolic blood pressure, systolic blood pressure, unit of measurement and body position.

The principal contributors to the development of the implementation layer are usually representatives from Clinical Data Management with expertise in the designing and building of CRF pages in the EDC system and in the design of

non-CRF data models. They are advised by representatives from the DSO with expertise in SDTM to ensure that the implementation layer contains all the information required and can be easily transformed to the data tabulation model. It is at this stage where the data elements previously defined are assigned metadata so that they can be included within either a CRF page and/or a non-CRF model. Following the stage's completion, the proposed information and implementation layer are reviewed by the GIS-GC Across Therapeutic Area Clinical Core Teams, principally the Data Collection Core Teams, and approved with appropriate feedback and escalation loops built in for the resolution of comments.

The final stages in the development and maintenance of data collection standards involves the creation of any edit or validation checks which can be included either within an EDC system or as part of an acceptance testing procedure to ensure that the data entered or received is as accurate as possible and is structurally as expected. Once again this stage is supported by representatives from Clinical Data Management with expertise in the building of CRF pages in the EDC system and in the design of non-CRF data models. These are also reviewed and once finalized the entire package is implemented in the relevant tools and they are adopted for use for all new studies within that therapeutic area or indication.

DATA TABULATION

The development and maintenance of the Global Data Tabulation Standards is driven by either an update to the Global Data Collection Standards (changes and additions) or by an introduction of a new version of the SDTM model. Both of these scenarios may drive an update to the annotations for the data elements contained in the Global Data Collection Standards and/or the Roche extensions to SDTM and, in the case of a new version of SDTM, an update to the data tabulation model itself.

Any update to the Global Data Tabulation Standards needs to be carefully considered for both the impact on the implementation and for the downstream impact on analysis programming. If the annotation of an existing data element is updated this also needs to be updated as part of the standard programming templates used to create SDTM datasets. As SDTM is the source for analysis datasets, any update to SDTM must be evaluation for any impact on standard analysis tools and programs as a change here may, in the worst case, require these to also be updated.

In the case where there has been an update to the Global Data Collection Standards a representative from the Data Tabulation Core Team will have been assigned as part of the development process. The annotations are created in parallel to the development of the implementation layer and are reviewed by the core team as part of the approval step within that process. Based on these new annotations an assessment is also carried out to ensure that the data tabulation model is aligned with any potential changes contained within the annotations.

The Data Tabulation Core Team review and provide comments on all proposed new SDTM versions as part of the CDISC public comment period. A preliminary impact assessment of the new SDTM version will also be performed at this time to allow for an initial communication to be shared with functions responsible for either creating or using SDTM. Following the final release of a new SDTM version the annotations of all data elements will be reviewed and updated as appropriate. For example, as part of SDTM IG v3.2 a sponsor defined interventions domain for procedures, XP, will be replaced by the new CDISC defined procedures domain, PR. The data tabulation model and extensions are also updated at the same time. All new annotations and updates to the data tabulation model are reviewed and approved by the Data Tabulation Core Team.

DATA ANALYSIS

The primary focus for the two Global Information Standards Data Analysis Core Teams is the development of new Global Data Analysis Standards. All technical specifications must be driven by an analysis concept and the process for the development of Global Data Analysis Standards reflects the expectation that no new technical specification can be approved without an aligned analysis concept.

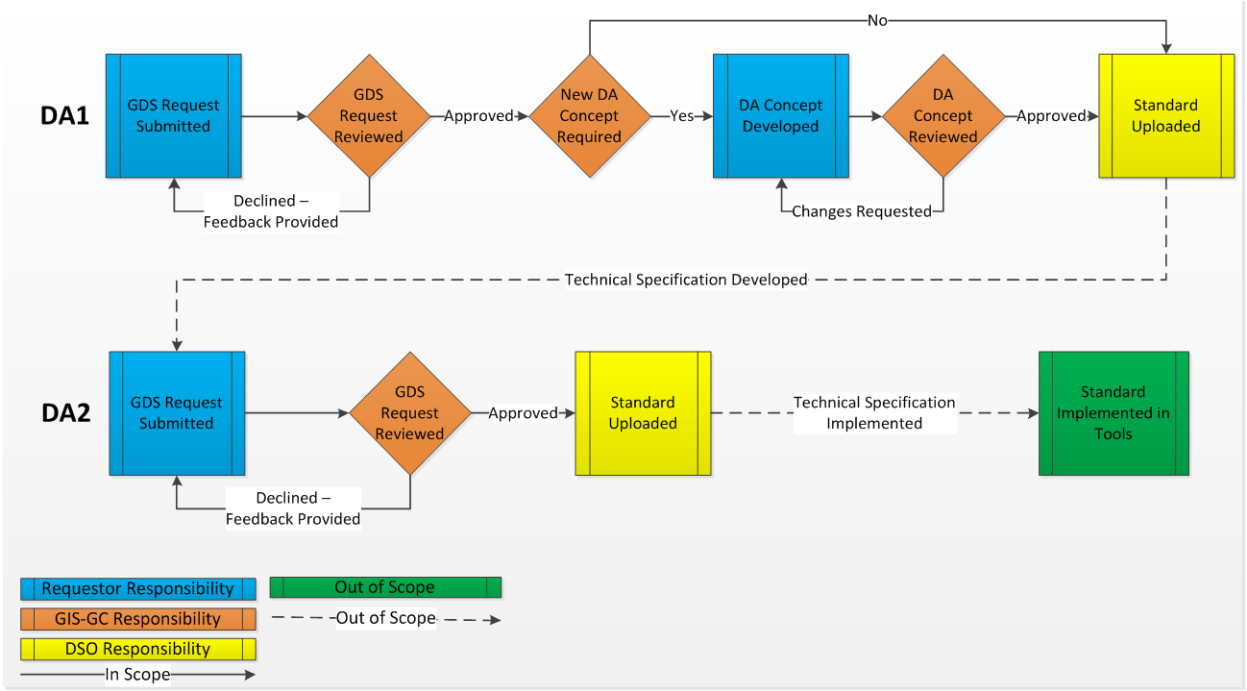

Figure 5: Global Data Analysis Standards Process Flow

The expectation is that all new Global Data Analysis Standards requests will originate with a request for a new table listing or graph. These are submitted by project teams for consideration to be included within the standards based on an assessment by the Project Lead Statistician as to the applicability across multiple studies, projects or therapeutic areas. Once a request is received, it is reviewed by the Data Analysis Definitions Core Team and either approved or rejected. Once a request is approved additional metadata is developed with the assistance of the submitter to define any required analysis or statistical concepts that are not currently contained within the Global Data Analysis Standards. Following the approval of these new analysis or statistical concepts by the Data Analysis Definitions Core Team they are then available for project and study teams to implement. Although, at this stage, no guidance on how this should be implemented from a technical perspective is available.

Once the definitions are finalized a separate team uses these as a basis to develop technical specifications for the table, figure or listing and any associated analysis dataset updates, from the addition of a new parameter to the creation of a new analysis dataset. These will then be reviewed by the Data Analysis Technical Specifications Core Team to ensure that they match the non-technical analysis concepts and, in the case of analysis datasets, to ensure that they are aligned with the Roche implementation of ADaM. Once approved these will once again be made available for use by both project and study teams as well for implementation within standard programming tools.

STUDY LEVEL IMPLEMENTATION OF GLOBAL DATA STANDARDS

While there is an expectation that all of the Global Data Standards are implemented as specified for every study, there is also an understanding that each study has the potential need to deviate from the defined standards. Each protocol is written for a particular purpose and may include some activities or biomedical concepts not included within the Global Data Standards. There are also examples where a protocol may not require some Global Data Standards described as mandatory (e.g. treatment pages on observational studies). The governance process has been developed to support study teams in managing these deviations (including both changes and additions) and to allow for these to be tracked and monitored. This allows for study teams to benefit from the experiences of other study teams by allowing for an institutional memory to develop and for potential necessary updates to the global data standards to be identified from the submitted requests.

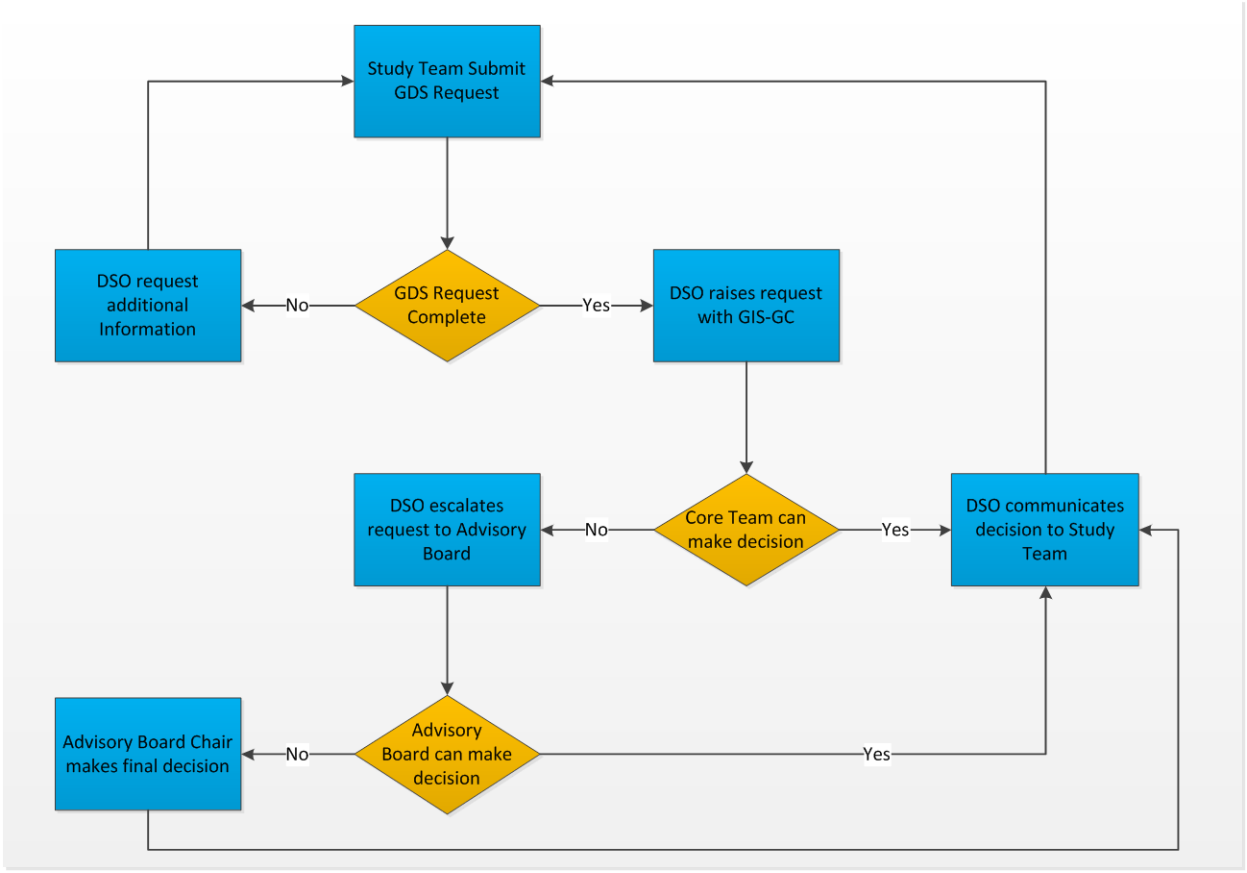

Figure 6: GDS Request Process Flow

The Global Data Standards request process, as defined, utilizes the GIS-GC to enable effective decision-making. Once a study team has submitted a request it is reviewed by a Global Data Standards Manager from the DSO. They will either request additional information from the study team to allow for an informed discussion or, if it is complete, raise the request with the appropriate GIS-GC Core Team. If a request has previously been raised and approved or declined then the Global Data Standards Manager is empowered to bypass the GIS-GC and make a decision based on precedence. At this stage Therapeutic Area Standards Experts and/or the Extended Team may be consulted for their input into the request. In the usual case the GIS-GC Core Team will make a decision and that will be communicated back to the study team. In the very rare cases that an agreement cannot be reached within the GIS-GC Core Team the request is escalated to the GIS-GC Advisory Board where the Chair of the Advisory Board is the final decision maker if necessary. In all cases the DSO communicates the decision back to the study team.

The request process is managed by a number of web forms which link to an issue management system to allow the Global Standards Managers to self-assign themselves to requests, to store all requests centrally, and for the tracking of metrics relating to number of requests received and response times. The use of web forms is an easy way to ensure that all required data is submitted, thus allowing for timely responses. In general, requests related to CRFs are responded to within 5 working days and those relating to non-CRF and SDTM annotations are responded to within 10 working days.

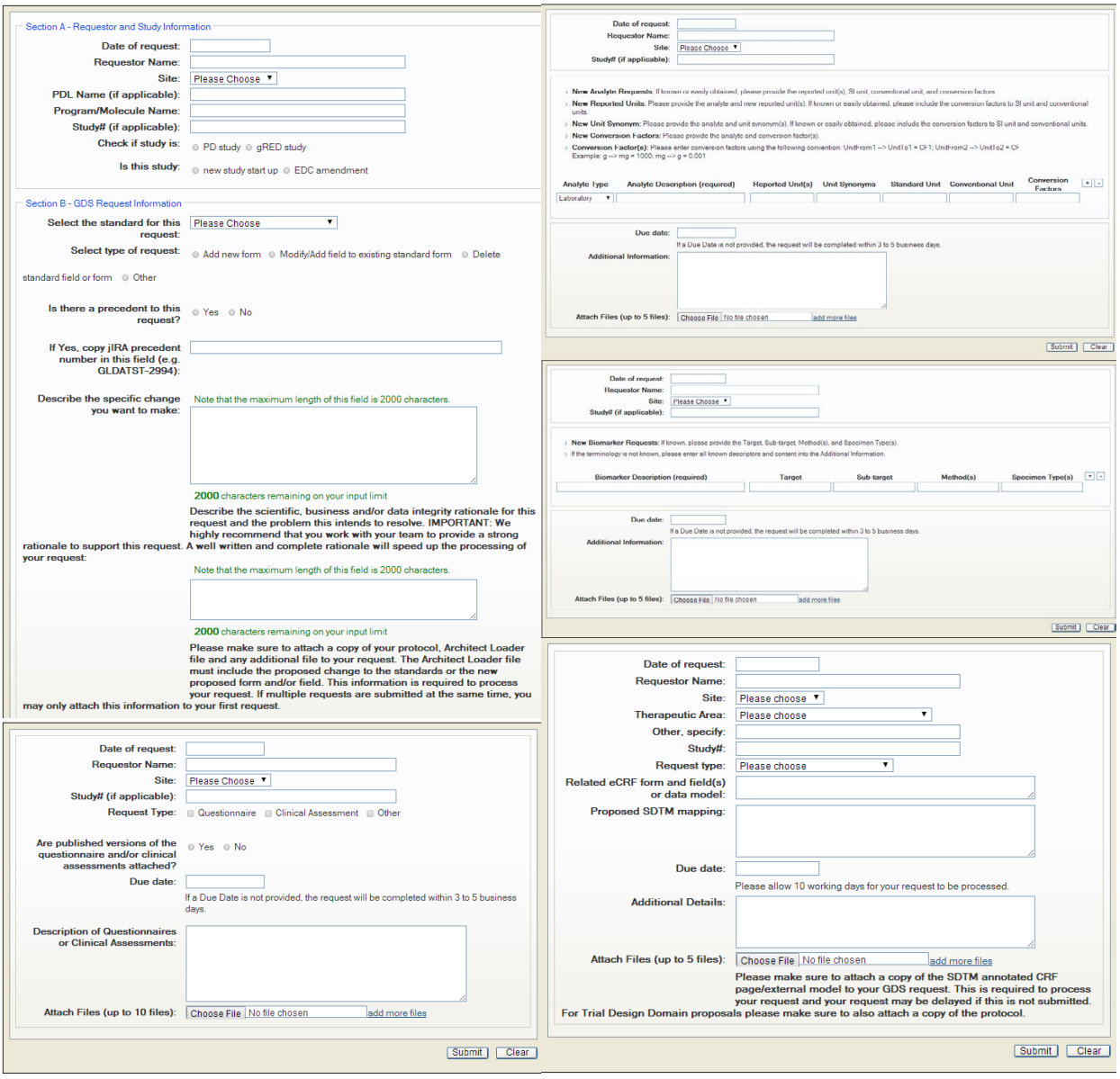

Figure 7: Global Data Standard Request Forms

CHALLENGES

While the process of implementing the Global Data Standards is well defined and supported there are a number of challenges in the practical implementation of it on individual studies. A key challenge is the lack of understanding, relating to the importance of the use of information standards, within the functional representatives in the study teams. This manifests itself in many ways and is best addressed by continuing education for all functions with responsibilities involved in the execution of clinical studies. A key requirement is to address the view that a particular study is "special". While a study may have particular elements which are innovative and novel, a highly significant portion of the study will be repeating elements from other studies, allowing for re-use. One particular area where this lack of understanding is particularly pronounced is in the absence of a protocol governance process which currently allows protocol authors to update the standard protocol template with unforeseen effects on downstream activities, in particular on data collection.

As requests relating to data collection are usually on the critical path of the study database build process there is a need to make decisions for study level requests as quickly as possible. It is therefore, sometimes, necessary to compromise between ensuring the "perfect" resolution of a request and the timeliness of the response. In these cases, a practical solution is sought and agreed with the study team so that the study build can progress. However, this would not be viewed as a precedence for other studies to implement. If appropriate, the proposal can be considered as part of the process of developing new standards. With the timeliness of responses being key, it is extremely important that complete information is provided as part of the initial request, as inevitably the need to return to the study team for additional information will have an impact.

Almost all large external vendors can, and do, provide study data in a structure with controlled terminology that matches the Global Data Standards. However, some smaller, more specialized, vendors have understandable difficulties in providing study data that matches the Global Data Standards. This can sometimes present significant challenges in the tabulation of the data provided, as additional programming is required to transform received data to the desired format.

DATA COLLECTION

All in-house studies (including both early and late stage protocols) are required to use the Global Data Collection Standards to collect safety and therapeutic specific information. Some flexibility is allowed, supported by a request process, for studies that need to deviate from the defined standards to meet a specific scientific or business need.

Study teams follow a pre-defined set of rules to identify instances where a request is required. For example, both the addition of a new data element, and the deletion or modification of a mandatory data element would require a request. Both a guidance document and an online training are available for study team members to guide them through the request process. Deviations from Global Data Standards must always be supported by a scientific or business rationale. To support this study teams are required to submit a copy of the protocol as well as any other relevant supporting documentation as part of the request to inform the decision making process within the Data Collection Core Teams.

The Data Collection Core Teams follow a set of principles when making decisions on study level requests. These include:

- Can existing Global Data Collection Standards be leveraged?
- Is this data collected elsewhere within the study?
- What are the proposed new elements required for? (e.g. data analysis, a specific safety reporting requirement)
- Is the proposal the best way to collect the requested data?
- How will this data be tabulated in SDTM?

DATA TABULATION

The implementation of Global Data Tabulation Standards within individual studies is mandatory and deviations from existing standards are not permitted with the exception of updates to variable lengths to avoid truncation. For data elements that are not part of the Global Data Collection Standards study teams are required to submit an annotation proposal for review by the Data Tabulation Core Team. These proposals are all reviewed and subsequently approved or revised with feedback provided to the study team. To guide study teams in the annotation of non-standard data elements a large number of formal guidances have been written to address the most common scenarios. These Guidance documents are regularly updated and the breadth of coverage increased based on the requests received from study teams.

DATA ANALYSIS

There is currently no defined governance for the study-level implementation of Global Data Analysis Standards, including analysis concepts, tables, figures and listings, and analysis datasets. There is a strong expectation that study teams use Global Data Analysis Standards where they exist, without deviation, if the analysis is required on a study. Where a deviation from a Global Data Analysis Standard is required on a study it is expected that the Project Lead Statistician notifies the GIS-GC for awareness. This maintains the flexibility of the application of the standards and allows the GIS-GC to monitor the implementation and to propose updates to the Global Data Standards if necessary. The GIS-GC, in rare cases, can also flag some deviations as not to be implemented in future studies where it is decided that the proposed deviation does not add any value to the existing standards (e.g. update to table structure without an update to the content).

CONTROLLED TERMINOLOGY

Where controlled terminology for assigned or derived values is not available for a study (e.g. new laboratory tests or PRO instruments) teams can request this metadata to be provided. This metadata is also included within machine-readable files, which are source files for template programs used in the creation of SDTM. The metadata matches a unique Roche identifier with the entire test driven SDTM variables for a particular domain (e.g. METHOD, SPEC, TEST, TESTCD). Also provided at this stage, for laboratory tests, are the parameter codes to be used in the analysis datasets, combining the multiple unique identifiers for a parameter in SDTM to a single identifier in ADaM.

FUTURE PLANS

A proposal to create an additional across therapeutic area clinical core team to the Global Information Standards Governance Committee, with responsibility for study protocols, was recently agreed. The Protocol Core Team will be responsible for harmonizing the development of standard protocol templates and content, and the enforcement of their use on individual studies. The inclusion of this additional core team in the Global Information Standards Governance Committee will allow for better coordination and synchronization between changes made in the protocol templates and those made in the Global Data Standards. Once established study teams will be required to use the mandatory elements of the protocol templates without deviation. It a study team requires to deviate from the protocol templates then a request will need to be submitted to the Protocol Core Team, with an appropriate scientific or business rationale for approval.

As technology improves, new opportunities in the development, management and use of standards become available. Roche's metadata registry (GDSR) is currently a store for all Global Data Standards Metadata and has the functionality for global metadata to be assigned a status of Proposed, Candidate, Standard, Superseded, and Retired.

- Proposed items have been requested by study teams and are approved or rejected for use in individual studies by the Global Information Governance Committee.
- Candidate items have been fully researched by the DSO, have sufficient metadata for use in a study, and are expected to be a standard in the near future and are available for use in all studies.
- Standard items have been fully researched by the DSO, have complete metadata, and are available for use for all studies.
- Superseded items have been replaced by a subsequent data element and are not available for use on studies.
- Retired items were previously proposed, candidate or standard items that are no longer considered appropriate for use and not available for use on studies.

Future enhancements within the Global Data Standards Repository will enable the storage of study metadata. The inclusion of study metadata within the Global Data Standards Repository allows for a more consistent end-to-end implementation of data standards. Each study will be able to select a subset of the Global Data Standards that are being utilized on that particular study and add proposed data standards directly to the Global Data Standards Repository. These proposed items can then be governed by the GIS-GC and either retired if rejected or included as study metadata if approved. This will also give the option for proposed items to be promoted to either candidates or standards based on the input from the GIS-GC.

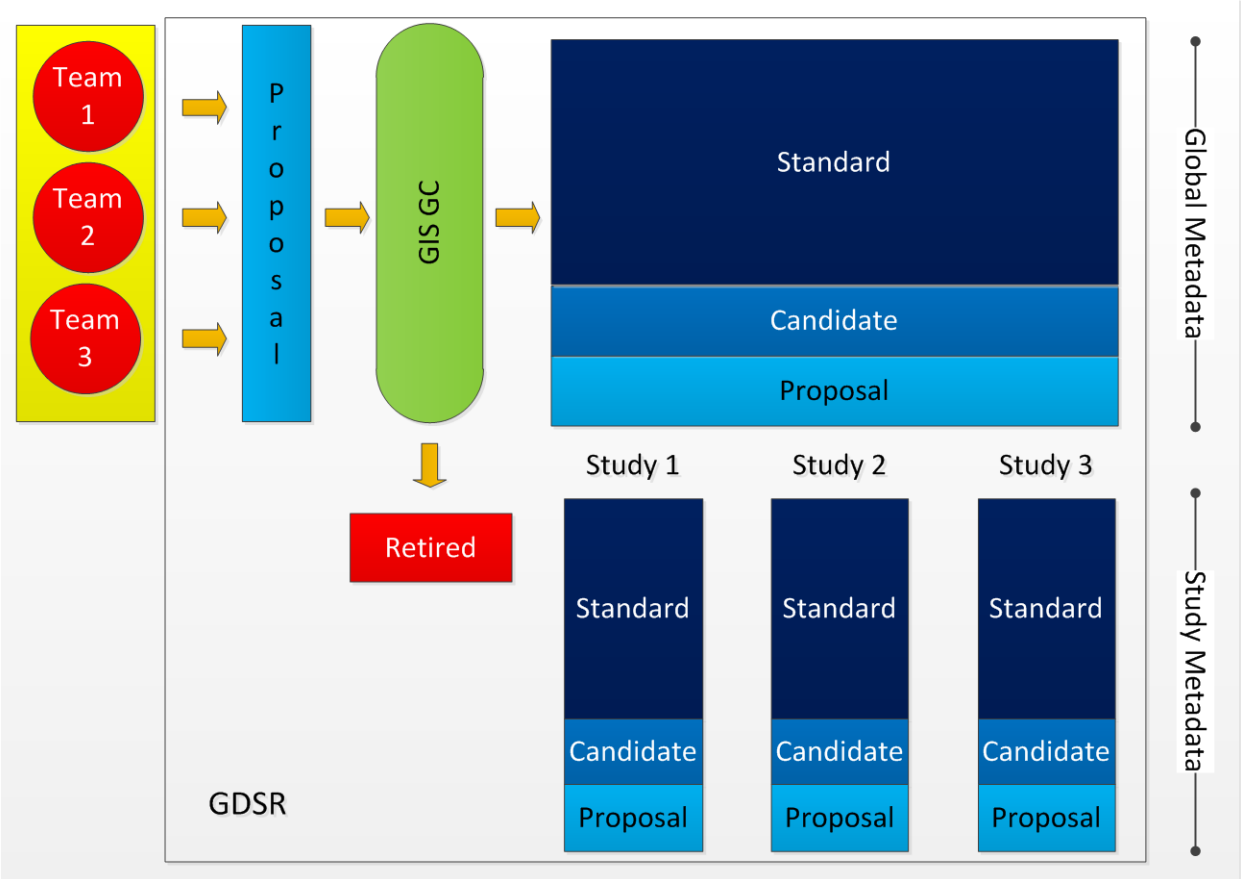

Figure 8: Future State of the Global Data Standards Repository

CONCLUSION

The most significant element relating to the successful implementation of data standards is an effective governance model. This must be able to support individual studies in their adopted of standards and be flexible enough to allow for justifiable differences between different protocols to be accommodated. By engaging and involving a cross-functional group of experts at varying levels of seniority within individual functions, Roche has been successful at managing the change from a more liberal governance model to today's formal model. As well as supporting individual studies the data standards must be continually developed and maintained so that they can adapt in an ever-changing scientific and business environment. This level of development, maintenance, and governance requires the support of both dedicated individuals and technology, especially on the scale it has been adopted at Roche.

ACKNOWLEDGMENTS

We would like to thank all the members of the DSO for their support in the development of this paper and the members of the GIS-GC past and present for their commitment to standards development at Roche.

CONTACT INFORMATION

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

Chris Price
Roche Products Ltd.
6 Falcon Way
Shire Park
Welwyn Garden City
Hertfordshire
AL7 1TW
Email: chris.price.cp1@roche.com