

ModSpace: Promoting Dissemination and Reuse of Models and Modeling Results

Jonathan Chard, Mango Solutions, Chippenham, UK
Henning Schmidt, Novartis, Basel, Switzerland

ABSTRACT

For small groups working in highly specialized fields, it is often efficient and sufficient to communicate knowledge among colleagues. As groups grow or as they get more important, mechanisms must be put into place to maintain and efficient use of the complex knowledge that accumulates. Here, we present one such a mechanism – the Drug and Disease Model Library – that is adapted to the work of the Modeling and Simulation line function at Novartis. The mechanism is simple, focuses on the dissemination and reuse of models and modeling results and utilizes a database to store and retrieve model based knowledge and techniques. The Drug and Disease Model Library project aims at providing a business process including a powerful and user-friendly software application called ModSpace. The ModSpace software provides a framework for storing, finding, accessing, modifying and re-using modeling information.

ModSpace is a collaborative effort by Mango Solutions and Novartis; this paper describes the software system that was developed, and how a user oriented focus will ensure that the system will be accepted and become part of the Novartis modeling activities on a day to day basis.

INTRODUCTION

It is widely accepted that in knowledge intensive domains such as the Modeling and Simulation (M&S) department of Novartis, productivity and efficiency critically depends on good knowledge management practices. As long as working groups are small, communication between colleagues is efficient and simple to maintain. When groups grow, as it was the case with the M&S line function of Novartis, other solutions must be found. The solution proposed here is to introduce an easily searchable repository of models, knowledge and experience to share the information among a larger number of group members.

The repository, called ModSpace will contain useful and working models and associated code, including documentation and keywords which might be relevant for more than one clinical/pre-clinical project. ModSpace is intended to be a living, updatable registry for ready-to-share models and code from internal and potentially also from external sources. It will primarily be used by associates within the Modeling & Simulation Line Function. Read access to the meta information, stored in ModSpace, will be provided to all Novartis Development associates. Its intended purpose is to leverage previously gained M&S project knowledge by:

- Facilitating internal sharing & updating of models and/or software code that may be relevant to more than one project
- Helping users to search and eventually find relevant entries.
- To present the search results in a way that supports the user in deciding of which available information to use as basis for his current modeling task.

This setup should ensure:

- Increased productivity and quality through reuse and extension of previous work.
- Reduced duplication of effort (not reinventing the wheel).
- Standardization - work evolves to a common format.

It is envisioned that ModSpace will fully become part of the M&S modeling activity workflow. It is intended to be the natural starting place for modelers when they begin work on a new data analysis question where they can:

PhUSE 2009

- Search for work that they have done in the past.
- Search for work (created internally and/or externally) other modelers have made available.
- Download, adopt, enhance and leverage previous work from colleagues and peers.

It is also intended to be the natural ending place for modelers when they complete their analysis work so that they may:

- Publish newly created ready-to-share models.
- Publish extended, ready-to-share versions of existing models.

PRIMARY USAGE SCENARIOS

SEARCH FOR MODELING INFORMATION AND POTENTIALLY DOWNLOAD:

1. Modeler receives a new “question to be answered” and initiates a search to check the ModSpace System for similar work that others may have been done in the past.
2. Modeler reviews information/content returned by the system. Determines whether information can be applied to their current modeling task.
3. If yes, modeler downloads content and may implement or adapt it to current modeling task.
4. System sets a “feedback notification” to inform modeler that a comment on usefulness of a downloaded model needs to be provided as a result of performing a download.
5. User provides comments and feedback on the usefulness of the downloaded information.
6. If no useful information/content returned, “start from scratch” approach to model development undertaken.

PUBLISH MODELING INFORMATION:

1. At the conclusion of a modeling activity a modeler is required to upload and publish the work in the system.
2. Modeler compiles all modeling information (model code, scripts, reports documentation, used data, and so on) and utilizes the entry/upload features provided by the system to publish the work.
3. As part of the publishing process, it is the responsibility of each individual modeler to determine what is a reasonable level of documentation and reusability for the models.
4. The system allows the owner to update and modify own entries to revise and improve the level of supporting documentation and modeling information for improved reusability.

INTEGRATION WITH MODELING ACTIVITIES

Modspace will be integrated with the standard workflow of a modelers activities. A project is not considered complete until it has been uploaded to the repository. Figure 1 gives an overview over the different types of possible uploads to ModSpace and defines which uploads are mandatory and which are optional.

PhUSE 2009

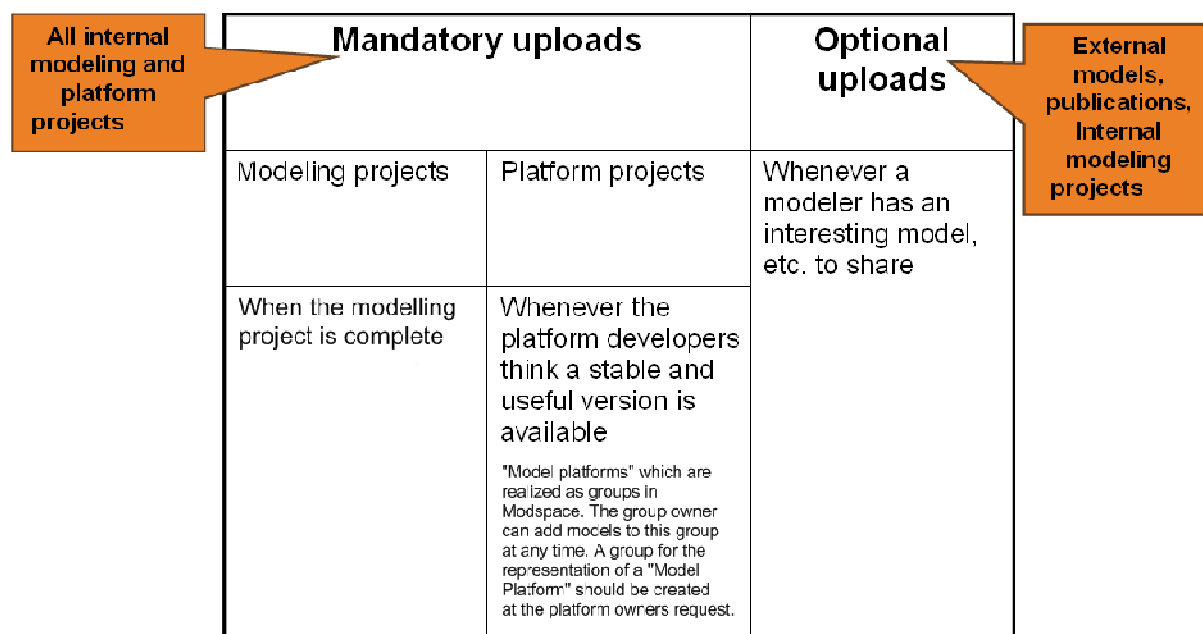


Figure 1: Different types of possible ModSpace entries.

"Modeling Projects" are modeling activities that answer specific questions arising in the drug development process. All such activities should be stored in ModSpace to serve as potential starting points for new projects. "Platform projects" refer to methodology and infrastructure improvements, such as computational toolboxes for certain tasks.

Users may also add models to the repository that they consider to be interesting; this may not be a model that they have developed as part of their work, but for instance could be some research that they have found which is interesting enough to be shared.

CREATING AN ENTRY IN MODSPACE

We use the term "Entry" to describe anything that a user may want to add to the library that they wish to share with others. This would be items such as:

1. A completed modeling project related that the user has been assigned. The modeling project is comprised of test data, scripts, outputs, supporting documentation explaining the model, and documentation that describes the model outputs
2. A template for a new type of model, with placeholders to be completed by the user that downloads it
3. Standard scripts and toolboxes that will be of use in other projects
4. External literature which describes methodologies, tools and techniques

Any of these can contain:

- documentation that describes how the model was developed and explaining how it can be used
- sample code
- test or sample data that will help another user to run the model
- model files and scripts to pre and post process the data
- documentation that explains the generated model outputs and supports the conclusions
- a folder structure, allowing to organize the different files in an entry

When a user has a collection of files, or a completed model that is ready to be added to the system, he can create a new entry. If required, the system can connect to Novartis source code and data repositories to extract all the files that the user has been working on. Otherwise, the user may upload all the files from their local workstation individually or all together in a zip file, that might include the desired folder structure. Each file that is added is processed, and necessary information is extracted.

PhUSE 2009

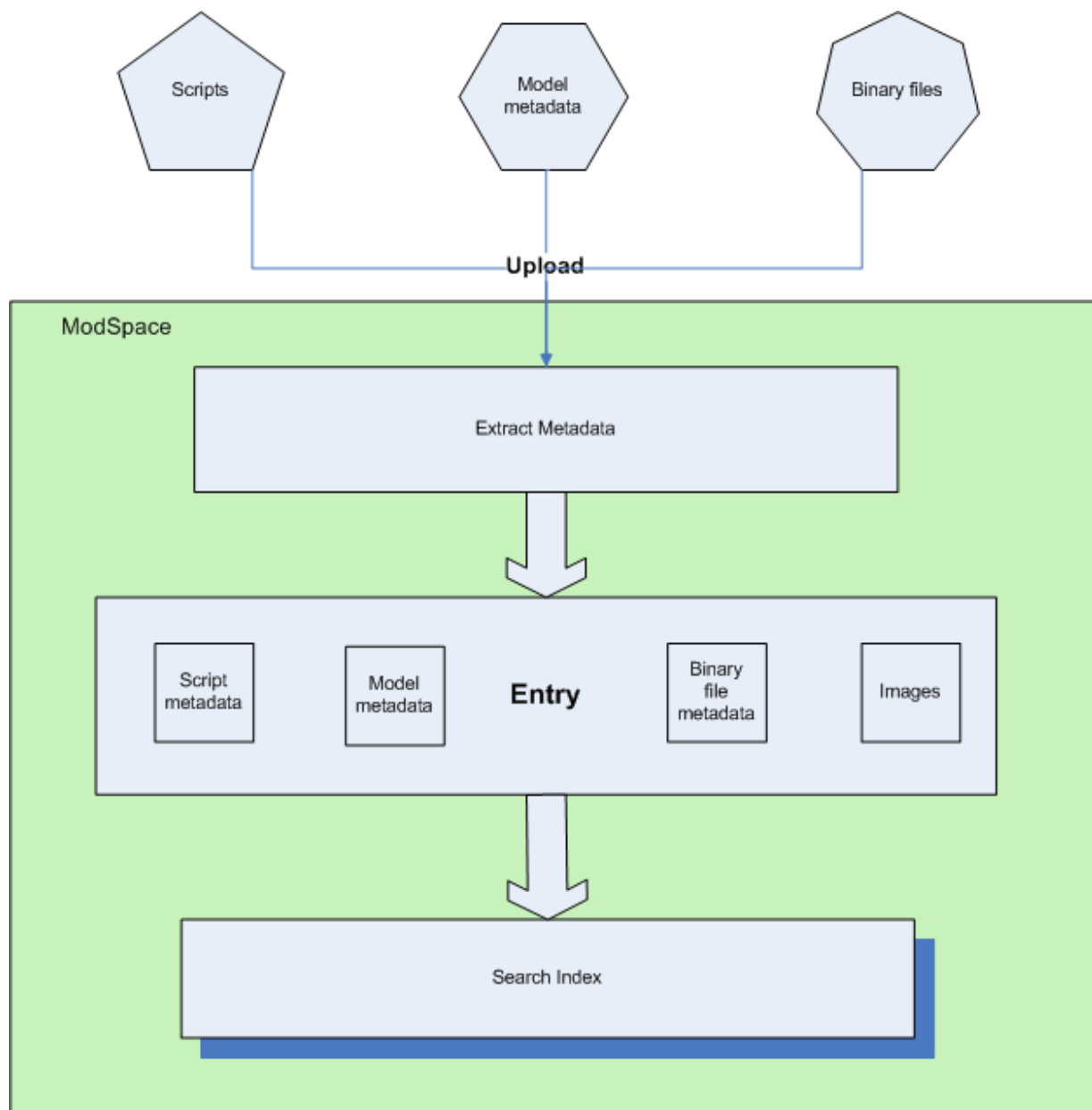


Figure 2: Uploading files

ENTRY METADATA

The user must set metadata on an entry which categorizes it and makes it easier for a user to find a solution to specific problem they have.

The user can set information such as (list not complete):

- Entry description
- The animal species to which this model applies
- The keywords associated with the model
- The therapeutic area
- The indication the model was developed for

This allows a user to home his search towards a relevant entry.

PhUSE 2009

FILES INCLUDED IN A MODSPACE ENTRY

The user can add data and script files (typically ascii files), binary files (such as PDFs and MS Office documents), and images to an entry. When a user adds a script file, two types of information are extracted from the files:

1. Text which is part of the script headers. If the user has conformed to the Novartis coding standards then the system will extract information such as the description, the assumptions, the inputs and outputs of the script, and sample code.
2. The system also understands the structure of the code that is added to the system. For instance, if an R file is added then the system extracts the libraries that the scripts required, and if a NONMEM control file is added then the system extracts information such as the problem statement, the parameters to be estimated (theta), and the covariance matrices (omega and sigma).

Currently the file types that are supported are R files, Matlab files, and NONMEM control files, although the system allows extending this functionality to other file types.

The system extracts all the text content from the files that are uploaded (including MS Office documents and PDFs) and adds it to the search index, so another user may search for anything that is contained within an entry.

The full text search capability of ModSpace, combined with the (higher ranked) search across the metadata of entries and files present in the entries is a powerful construct that allows the user of ModSpace to quickly find entries and files that are of interest for the modeling question to be addressed.

RELATED ENTRIES

An author may also indicate that the entry he has added is related to another entry; for instance, if he downloaded another entry and used it as the starting point of his work, or if another entry contains a particularly useful script that the user included in his entry.

This allows the system to maintain a genealogy of entries, and exposes entries that are particularly “inspirational”.

MULTIPLE ENTRY AUTHORS

The system allows collaboration on entries by supporting multiple authors on an entry. When an entry is created, then the originator of the entry can add “authors” to it. Other authors can also modify the entry metadata, and add and remove files from the entry.

PUBLISHING ENTRIES

When the user is satisfied that all the files have been added to the entry, and that the documentation is in place, the user should publish the entry. The user may also modify an entry that he has previously published. When the user publishes an entry, he enters a comment to describe changes he has made. This allows the user to make it clear that a new version of the entry resolves some issue, or contains an enhancement.

Publishing the entry causes the system to store the entry metadata and the entry contents in the search index, which then makes it available to other users.

GROUPS

A group is a collection of entries that share a certain theme, or address a particular problem. The owner of a group can add other entries to the group, and manage the structure of the group – the group can be arranged into folders that contain a set of entries.

A group also contains a set of metadata that can be used to describe, summarize, and make the group appear in search results.

SEARCHING FOR ENTRIES

As noted above, a user can search for entries that will aid the solution to his problem. The system supports a sophisticated (Google-like) syntax which allows the user to combine searches on any pieces of metadata. For example, the user can write a query which searches for entries that:

- Contain a certain keyword
- Were developed to address a certain disease
- Contain code in a particular language
- Contain an R script which uses a particular library (for instance, to find some sample code)
- Does not apply to a certain therapeutic area

PhUSE 2009

- Contains a file with a particular phrase

Each of these search fields can be OR'ed or AND'ed together.

VIEWING ENTRIES

Once a user has found an entry that is of interest, he can open it and view the files that are in the entry.

The system allows the user to view a preview of each script file that has been added, and view the metadata that has been extracted from/attached to each file.

If the user decides that the entry will be of use, then it can be downloaded to the users local machine and the model code can then be repurposed for the specific problem that the user is trying to solve.

When a user downloads an entry, the system generates a reminder for the user asking him to provide some feedback to the author regarding the quality of the entry. This allows the user to indicate that the entry was well documented, robust, was easy to reuse, and whether it saved the user time (and how much time it saved them).

The user can set a free text comment which is visible to all, to add more information about the model and what they thought about it. The user may also send a private message which will only be visible to the author, perhaps to suggest a possible future enhancement, or to report a bug.

The system also shows the history of the entry, indicating when it was updated and why, and listing all the feedback that has been submitted by other users. This allows a user to make an assessment of the entry, for instance if other users thought it was useful, and if the entry has been updated at all, for instance to include enhancements or fixes.

Finally, if the user thinks that the entry is worth "watching", then he can bookmark it for future reference.

HOME PAGE

When a user logs in, he is presented with a personalized home page containing information specific to that user.

The user can view:

1. Bookmarks
2. News
3. Feedback
4. Reports

BOOKMARKS

The home page displays all the entries that have been bookmarked by the user. The bookmark list displays all the entries that the user has created, and all the groups of which he is an author.

The bookmark list also highlights any entries that have been created or updated, but are not yet published.

The user can use the bookmark list to access all the entries that he is interested in.

NEWS FEED

The news feed for the user contains all the events that have been generated on entries that are in that users bookmark list. This shows the user when:

- An entry that he has bookmarked has been updated
- Someone has commented on an entry of which he is an author
- Another user has added a brand new entry

This allows the user to keep abreast of any developments on entries that he has shown an interest in.

FEEDBACK

The user's home page displays all the feedback that he has submitted to the system, along with any outstanding feedback that he has not yet provided. This allows him to easily submit feedback that is outstanding once he has assessed the usefulness of models that he has downloaded. He may also modify feedback that he has already left – for instance to add or update the comment, or indicate how much time the model saved him.

PhUSE 2009

REPORTS

A user can add reports to his home page, to show a summary of the state of the system. The following is an example of the types of graphic that the user can add:

- A list of the most recently added entries
- A list of entries that have been downloaded the most times
- Which entries have been given a the highest number of positive feedbacks
- Which entries have saved other users most time

Each of these reports can be configured, for instance to only show entries that are of a certain type, or which entries pertain to a particular drug type.

This allows the user to view a snapshot of state of the system, perhaps to keep an eye on entries which have been receiving acclaim from other users, or keep track of their own entries.

The reports will also be used by management to keep track of the benefits of the system – by assessing how much time the system has saved, and that quantity and quality of entries that are stored. Reports can be configured by team or drug type, so it is possible to compare and contrast sections of the community and perhaps identify areas of weakness – for instance, if there are few models that deal with a certain problem area.

DEVELOPMENT PROCESS

AGILE AND ITERATIVE

The system was developed in an aggressive time frame, and while the requirements of the system were well understood at the outset, the implementation and solution required were not.

Mango and Novartis worked together closely throughout the development of the system to ensure that the solution that was developed fitted the users needs. This was achieved by adopting an agile process which involved:

- Early development of a working system which addressed some aspect of the problem
- Regular presentations and demonstrations of the working system to key stakeholders in the project
- A “just in time” attitude to functional specification, which allowed development to begin on aspects of the system which were agreed on, while specifications of other parts of the system were being investigated and written

The first phases of the system were hosted internally on Mango's development servers, and presentations were made to Novartis to demonstrate the functionality which had been developed during that iteration. Later iterations were hosted on Novartis hardware, to test integration with the internal IT systems, for example how files were imported from the various source code repositories that are available.

The documentation surrounding the system is stored in a Wiki hosted at Mango. The Wiki stores all the information collected while each “User Story”¹ was written. Mango gave Novartis access to the Wiki so that they could view and comment on the information that was stored there. The Wiki is also configured to allow User Stories to be viewed in a list, sorted by status, anticipated delivery phase, and requirement. This allows both Mango and Novartis to view the most up-to-date version of all the functional specifications, and monitor the progress of the project.

INVOLVE USERS

It was vitally important that the future users were involved in the development of the system, as ModSpace will form an important part of the M&S activities within the drug development process when it is completed.

To that end, the presentation of the system to the users at the end of each iteration serves the following purposes:

- It gives the users a forum in which they can comment on the functionality, raise concerns, and react positively to things that they like
- It increases the profile of the project among the users
- It increases the familiarity of the users with the system so they are not “shocked” when they are asked to use it for the first time
- It mitigates the negative feelings that some users may feel when they are asked to use a new system.

INVOLVING USERS – APPLIED

A fine example of how this “early demonstration” of functionality affected the system design came when we

PhUSE 2009

demonstrated to the users how the “feedback” system would work.

It was originally envisaged that the users would be able to “rate” an entry – giving several aspects of the entry a rating on a sliding scale. When this was demonstrated to the users, we received a significant amount of negative feedback as it was considered that a poor rating would potentially be damaging to the entry author.

We reviewed the process of adding feedback to the system so that users could only submit either “no comment”, or a “positive rating” – for instance that the user thought that the documentation was good.

We demonstrated this to the users in the next presentation and the users much preferred this solution.

This type of process aids acceptance of a new system, and reinforces the view that the system was developed “for” the users, and that it is “their” system.

A second example of the advantage of demonstrating the system to users came in the very first iteration. This version demonstrated parsing of files that were added to the system, and extracting the metadata that was defined within those files. A user responded “Now there is a reason for following the coding guidelines!” This positive reaction enhances the systems reputation and validates its existence.

FUTURE DEVELOPMENT

We have designed the system such that it is easy to extend and add new types of files that can be parsed. The initial system supports:

- Matlab files
- R files
- NONMEM control files

We have plans later to add support for:

- S-PLUS files
- NONMEM output files

The tool also has a potential social networking aspect, which can be explored. At present the tool is relatively passive; a user views his home page and sees updates to entries that he has bookmarked. Depending on usage patterns, we could extend the system so that it becomes a messaging platform:

- A user submits a request to the community for a certain type of model
- A user finds an entry that can be of use, but needs some help with it, and sends a message asking for assistance
- The addition of forums or message boards to groups or entries

We can also add more reports to expose how the system is being used and demonstrate its benefits to the business.

CONCLUSION

The system is due to go live on the 15th December. The value of a system such as this is dependent on the quantity and quality of information that it contains, so it will take some time for the system to realize its true potential. There is a legacy of models and information which will need to be added to the system, and each of these models may need some attention before they are in a state where they can be shared with others.

However, it is anticipated that as the system is integrated with the daily activities of modelers, and its usage increases, that the wealth of knowledge and experience at Novartis will be made available across the entire organization. This will enhance the efficiency and productivity of the modeling and simulation teams, and avoid modelers duplicating work that has already been done by others. Rather, a library of useful and extendable models can be shared, which are continuously updated and improved, leading to a faster turnaround in modeling projects, growing the modeling community.

REFERENCES

- 1: [Definition of a user story by Agile Modeling](http://www.agilemodeling.com/artifacts/userStory.htm). <http://www.agilemodeling.com/artifacts/userStory.htm>

PhUSE 2009

CONTACT INFORMATION

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

Jonathan Chard
Mango Business Solutions
Unit 2
Greenways Business Park
Chippenham
SN15 1BN
Work Phone: +44 (0)1249 76700
Email: jchard@mango-solutions.com
Web: www.mango-solutions.com

Brand and product names are trademarks of their respective companies.