

# ModSpace: Analytical Knowledge Management

Richard Pugh, Managing Director  
[rich@mango-solutions.com](mailto:rich@mango-solutions.com)  
4th May 2011



# Agenda

- Mango Solutions
- Analytical Knowledge
- ModSpace
  - The Project
  - The Application
  - Technical Details
  - Wider Applicability
  - The Alternatives
  - The Development Path
- Summary & Questions

# Mango Solutions

## Mango Solutions

- Private Company founded in 2002
- Headquartered in the UK
- Offices in Switzerland, USA and China
- Global Team of 38
- Strong year-on-year growth since 2002

# Mango Solutions



## People

- Analytical Expertise
- Project Managers
- Business Analysis
- Technical Architects
- Developers
- Dedicated Testers
- Quality Manager

## Skills

- 'R' / S+
- Matlab, Python
- SAS
- Java / C++
- Oracle
- Web Reporting

## Services

- Training
- Commercial Support
- Technical Consulting
- Business Consulting
- Software Development
- Software Development

## Products

- ModSpace
- Navigator
- Push2Doc
- iNCAS
- ValidR

## Mango in M&S

- Work with M&S Groups from most top 20 pharma companies
- Provide training, consulting, support and application development services
- Also participate in cross-company projects such as IMI initiative



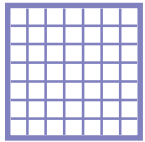
# Analytical Knowledge

# Analysis

- Analysis is the practice of producing a model (or “rule of thumb”) to describe a set of data
- We need to understand
  - How well a model “fits” the data?
  - How accurately a model performs?
  - What variability we can expect in our answers?



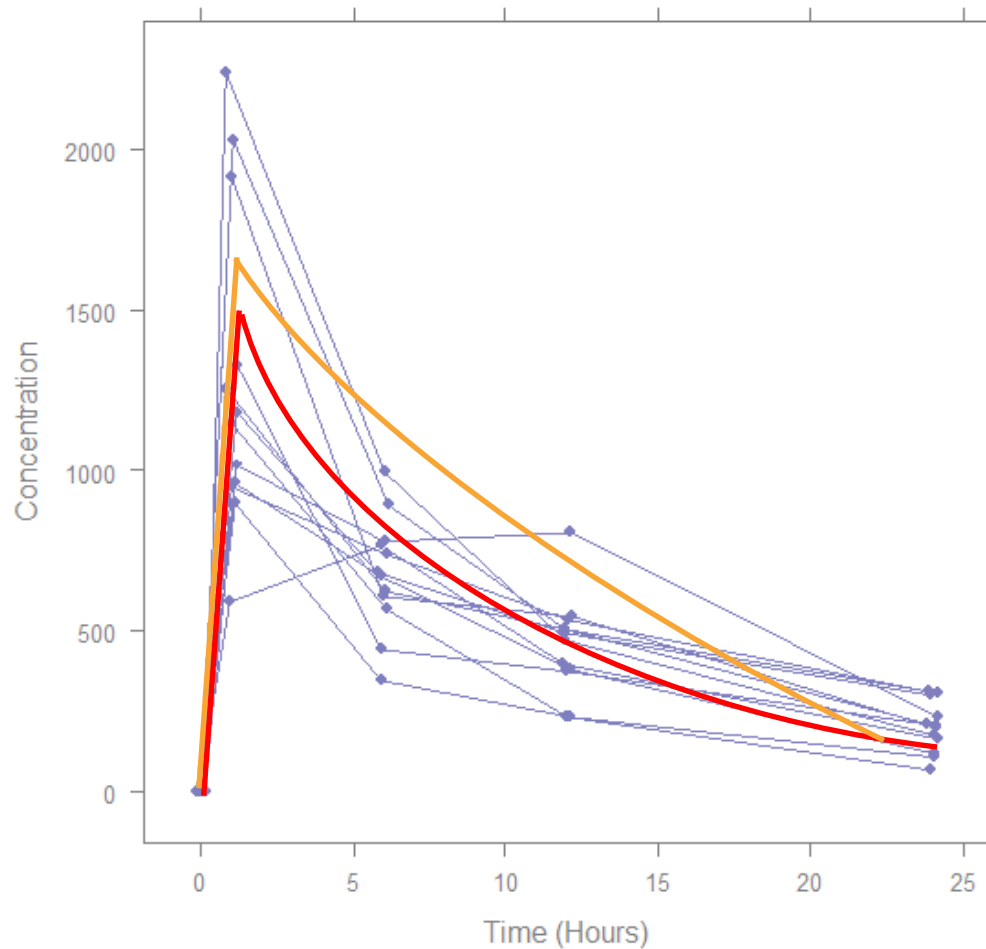
## PK Concentration over Time



Data



Models



Model Outputs

# Analytical Knowledge

- Contained in
  - Datasets
  - Analytical Programming Scripts
  - Graphics, textual and tabular outputs
- Models described mathematically in different analytical languages, and split across a set of files

## Why Store Analytical Knowledge

- More decisions being made more quickly on more data
- Analytic IP often difficult to reuse both by other analysts and beyond
- Difficult to clarify “what exists” before wheels are recreated, often not helped by typical “analytic” reporting lines

## Challenges for AKM

- Analytical Knowledge is typically complex and highly valuable
- Variety of analytical languages used
  - SAS - Large Corporation
  - R - Open Source Language
- Analysts are primarily not programmers
  - Often no coding standards
  - Little use of versioning

# ModSpace The Project

## ModSpace Project

- Part of Technical Mango-Novartis Partnership
- Agile Software Development project
- Part of Novartis' "MODSIM" platform (more later!)
- Project Timelines
  - Initial PoC in January 2009
  - Initial URS May 2009
  - Agile Implementation from May to Nov 2009
  - Into Production December 2009

## ModSpace Project

*“we should write whitepapers using this as an example of how a software development project should be run”*

- Good initial PoC with visual design outputs
- Agile Development with Fixed Scope
- Very strong input from the business
- Excellent working relationship

## ModSpace Project

- Project Aims
  - Central storage and description of “models”
  - Easy to find and download models
  - Feedback mechanism
  - Add versioning to analytical files
  - Encourage use of coding standards
- Initial Project name “Moogole” hints at vision!



## Design Concepts

- Allow description of a set of files as a “model”
- Storage of different file types
- File-type-specific parsers to extract as much meta data as possible based on file structure
- Experience based on social media applications

## Project Outcomes

- Big success story within the Mango-Novartis Technical Partnership
- Lots of Good Information being stored
- Some unexpected uses (e.g. storing videos of training courses)
- Some challenges ahead around curation as more complex element types are stored

# ModSpace The Application

## Some Terminology

- Element - A Single File
- Entry - A Set of Files that, together, form a Group of Files that someone may want to find (e.g. a “model”)

# Application Workflow

- Add & Describe Information
- Collaborate with other Analysts
- Publish Information to the wider group
- Search for Information
- Standardised view of Information
- Download Information
- Provide Feedback on Information
- Create Communities
- Produce Management Reports

## Add Information

- Upload Files and Directories OR link to existing Version Control Repository
- Type-specific parsers and storage
- Parsers can encourage or enforce coding standards
- Creates entry in version control engine

# Add Information



Identify File Type

Parse and Extract Element Meta

Store Elements

Create and Describe Entry

## Collaborate

- Initially, the files are hidden from view
- Can add members to the “Entry” to collaborate on files before publishing to wider group
- Members can be anywhere on network (e.g. different countries)



# Publish

- Tags the “entry” and adds a “commit” comment
- News item automatically generated
- Added to the general news feed for users



Updated Entry: NONMEM 7 Training course from Uppsala University

Training Materials from NONMEM 7 Course taught by Mats Karlsson

*Updated to include additional NONMEM files used in the course*

*Richard Pugh, 20-Feb-11 08:43PM*



New Entry: Papers from ACOP 2011

Collection of papers from ACOP 2011

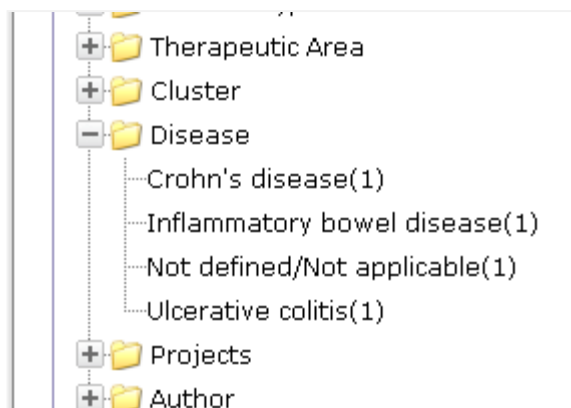
*Just got back from ACOP 2011 - some really good papers*


*Richard Pugh, 20-Feb-11 08:42PM*

## Search for Information

- Apache Lucene search engine behind the scenes:
  - Simple search
  - Advanced search
  - Google-Syntax search
  - Filtering of Results
  - Suggestions and Spelling-Matching

# Search for Information



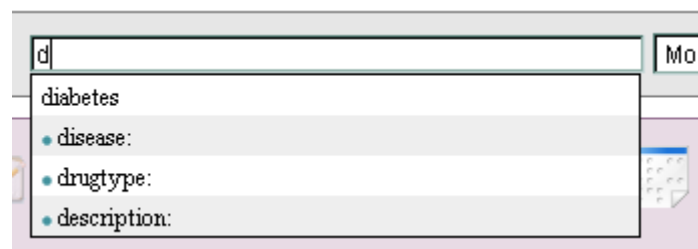
Name	Last Modified	Sc
 <b><u>Calculating individual weighted residuals correctly with different error models</u></b> This document describes how to code a NONMEM model in order to be able to calculate individual weighted residuals correctly. The document includes code examples and explanations for NONMEM only, but ... <a href="#">IWRES</a> , <a href="#">individual weighted residuals</a> , <a href="#">NLME</a> , <a href="#">nonlinear mixed effects</a> , <a href="#">NONMEM</a> , <a href="#">MONOLIX</a> , <a href="#">error models</a> <i>Liam Shields , Helene Karcher , Roland Fisch</i>	05 Oct 2009	
<b><u>CXCL10 target summit</u></b> This folder contains MATLAB code and associated documentation used for target summit modeling		

## Search Results

No results found  
Your search **pratform** did not match

**Did you mean?**

[platform](#)  
[perform](#)  
[reform](#)



# Standardised View of Information

- Each Entry has the same initial “view” to allow easy analysis of applicability
- Each element has type-specific views
  - Single page Meta Description
  - Syntax-Highlighted File Preview
  - History view of Versioning

# Standardised View of Information

File
 Folder
 Delete

Edit
 Related
 Owners
 Publish
 Download
 Delete
 Group

**Entry Contents**

- One Compartment PK Analysis
  - Sling Dog
    - Additive
      - output
        - additive.nmctl**
    - Constant CV
      - constant\_cv.nmctl
    - Log Normal
      - log-normal.nmctl

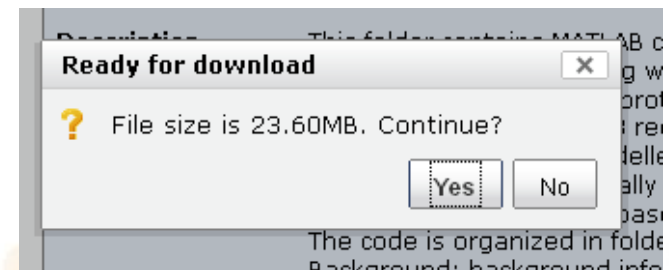
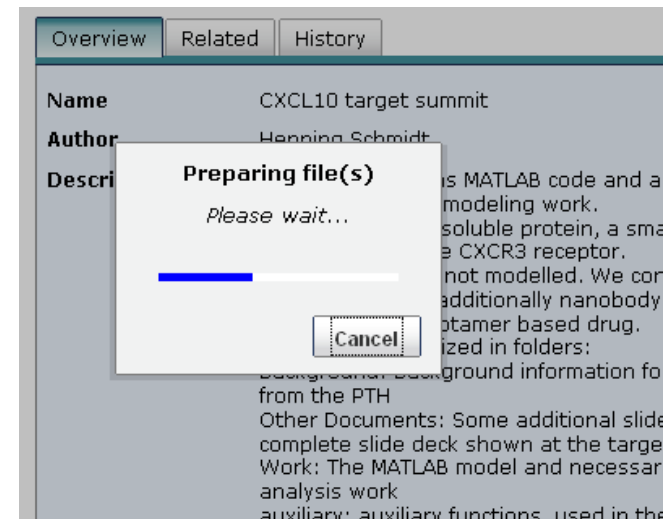
Overview | **File Preview** | History

```

$PROBLEM Fitting a two compartment model with constant infusion
$INPUT C ID AMT RATE TIME CT=DV LCT PD1 PD2 EVID
$DATA DOG1.CSV IGNORE=C
$SUBROUTINE ADVAN3 TRANS4
$PK
  CL=EXP(THETA(1))
  V1=EXP(THETA(2))
  Q=EXP(THETA(3))
  V2=EXP(THETA(4))
  S1=V1
  S2=V2
$ERROR
  Y=F+ERR(1)
$THETA
  -3
  0.1
  -4
  0.1
$OMEGA
  800
$ESTIMATION MAXEVAL=9999 PRINT=1 NOABORT
$COVARIANCE
$TABLE NOPRINT FILE=inf2.tab ONEHEADER C ID TIME RATE AMT LCT PQ V2
          
```

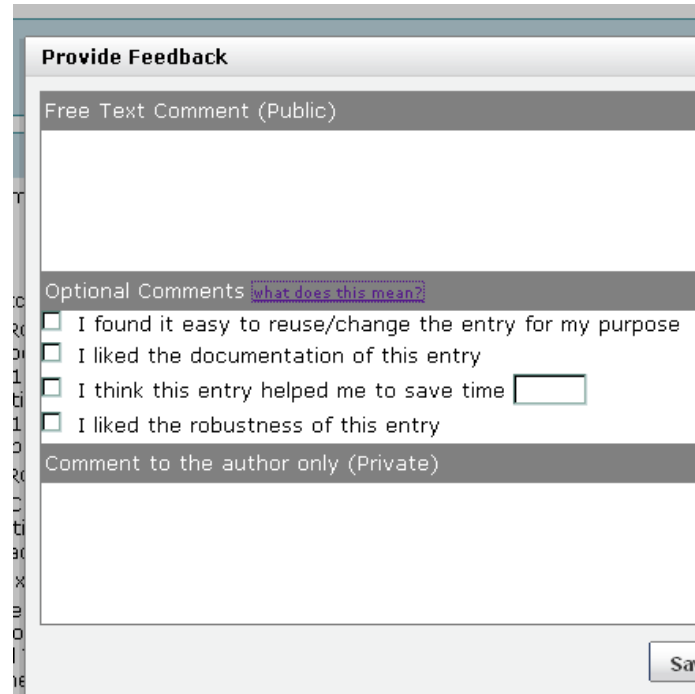
## Download Information

- Download single “Elements” or entire “Entry”
- Extract as Zip or work directly with version control repo
- Entry is “Bookmarked” and “Feedback” event triggered

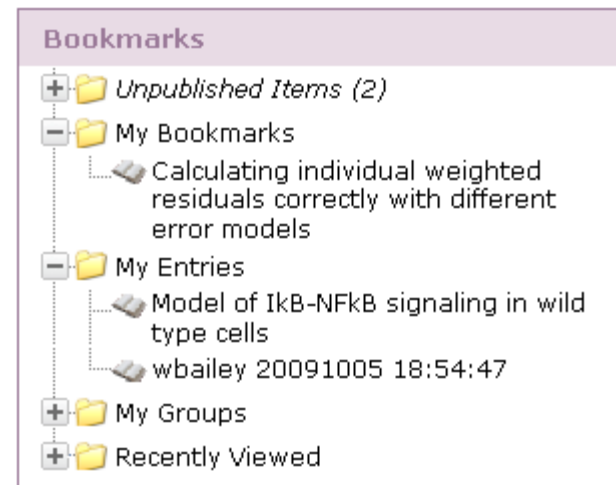


# Provide Feedback

- Feedback allows users to rate/comment on entries
- Provides feedback mechanism for bug fixes
- Feedback Information available for Management Reports



The screenshot shows a web form titled "Provide Feedback". It has two main sections: "Free Text Comment (Public)" and "Optional Comments". The "Optional Comments" section contains four checkboxes with corresponding text: "I found it easy to reuse/change the entry for my purpose", "I liked the documentation of this entry", "I think this entry helped me to save time" (with a small input field), and "I liked the robustness of this entry". Below these is a section for "Comment to the author only (Private)". A "Save" button is visible at the bottom right.



# Create Communities

- Create “Groups”, a collection of bookmarks within a specific category
- Has it’s own membership list and metadata

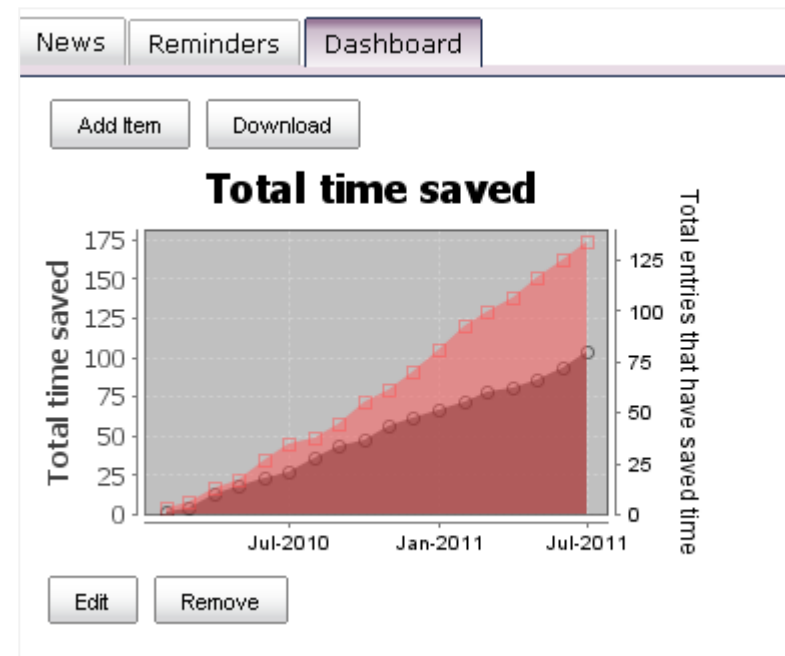
The screenshot displays the 'Group Contents' page for a group named 'R scripts'. The page is divided into two main sections. The left section, titled 'Group Contents', shows a tree view of the group's contents: 'R scripts' (selected), 'Analysis scripts', 'R simulation package for M&S course for partners', and 'Graphics scripts'. The right section, titled 'Overview', contains a table of metadata for the 'R scripts' group.

Overview	
<b>Name</b>	R scripts
<b>Description</b>	A collection of R scripts for gen
<b>Keywords</b>	boxplot, tornado plot
<b>Therapeutic Area</b>	Not applicable
<b>Drug Type</b>	Not applicable
<b>Target Categories</b>	Not applicable
<b>Cluster</b>	Cardiovascular Medicine
<b>Language</b>	R



# Produce Management Reports

- Run Reports on Stored Meta Information and Feedback
- Create Standard Dashboards to assess value of Stored Information by User, Department etc



# ModSpace

## The Technical Details

# Technical Details

- Web-based Java application
- Apache Lucene Search Engine
- Hibernate Data Layer so Database Agnostic
- Interfaces with LDAP, PAM etc for security
- Easy to Administer



Manage Users



Manage Data  
Dictionary



Configure  
Elements



Configure  
Analysers



Audit Trail

# ModSpace Wider Applicability

## Wider Application of Software



- The “Recognised” elements can be extended and modified
- Version Control is enforced without user Knowledge
- Coding Standards can be encouraged OR enforced
- Feedback can be informal OR formal (i.e. peer review)

## Wider Application of Software

- ModSpace customers and prospects include:
  - The Bank of England for model management
  - An Insurance company for building communities for open source softwares
  - A pharma company who wants to create a “SAS Code Repository and Community”

# ModSpace

## The Alternatives

## Put your files on a Central Server

- Limited search
- No way to describe a “set of files” as a single “thing”
- No versioning and file management
- No intuitive interface
- No encouragement of standards and best practices



## Use Sharepoint

- Limited search
- Doesn't distinguish between “a script” and “a document”
- No way to describe a “set of files” as a single “thing”
- No encouragement of standards and best practices

## Use a Version Control System

- Limited search
- No way to describe a “set of files” as a single “thing”
- No intuitive interface
- No encouragement of standards and best practices

# ModSpace

## The Development Path

# Formal Entry Structure

- Enforcement of Project (File/Directory) Structure
- Validates Project Structure to Enforce Best Practices
  - Directory Structure and Naming
  - Existence of Files
- Additional Meta can be associated that Extends standard set of meta
  - Project Identification Number
  - Project Manager

# Discussion Groups

- What happens if you don't find what you're looking for?
  - Search online
  - Send an email to "all@"
- Adding Q&A feature so the question and answer are stored and searchable

## Questions



How do I calculate the risk of an asset within R?

Posted by rpugh on 1st Feb 2011

[See 8 responses](#)



Does anyone have a script to create a boxplot using the mean as the mid-point of the box instead of the median?

Posted by mcreed on 10th Jan 2011

[See 4 responses](#)

## Storage Types

- Currently, elements are stored in Version Control OR on file system based on type (data vs script)
- Can be extended so (for example):
  - Documents Stored in SharePoint or Documentum
  - Data Stored in Database

# Synchronisation with Version Control

- Can now:
  - Create an Entry in ModSpace
  - Connect to it via IDE (e.g. Eclipse)
  - Edit Files within Eclipse
  - See “needs synchronisation” message in ModSpace
  - Sync the Files
- Allows for programming Users and “Web” Users to work on same project

# Storage in Tech-Agnostic Format

- Part of the “ddmore” Initiative (part of “IMI”)
- Proposed Workflow
  - Analyst A codes model in Language A
  - Code checked into ModSpace
  - Code stored as an “Implementation” of the Model, which is also stored in a general format
  - Analyst B downloads the Model in Language B, adapts the Model and checks in code
  - Analyst A sees changes reflected in Language A



# Summary

## Summary

- Analytical Knowledge is typically split across files which need to be stored and described together
- The ModSpace Project was a bespoke software development project with Novartis
- ModSpace provides a general platform for analytical knowledge management

# Demo Modspace