ModSpace: Analytical Knowledge Management

Richard Pugh, Managing Director
rich@mango-solutions.com
4th May 2011

MANGOSOLUTIONS
data analysis that delivers

MANGOSOLUTIONS
data analysis that delivers

Agenda

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

MANGOSOLUTIONS
data analysis that delivers

Mango Solutions

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

Mango Solutions

ISO 9001

People	Skills	Services	Products
Analytical Expertise	'R'/S+	Training	ModSpace
Project Managers	Matlab, Python	Commercial Support	Navigator
Business Analysis	SAS	Technical Consulting	Push2Doc
Technical Architects	Java/C++	Business Consulting	iNCAS
Developers	Oracle	Software Development	ValidR
Dedicated Testers	Web Reporting	Software Development	
Quality Manager			

MANGOSOLUTIONS
data analysis that delivers

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

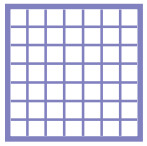
MANGOSOLUTIONS
data analysis that delivers

Analytical Knowledge

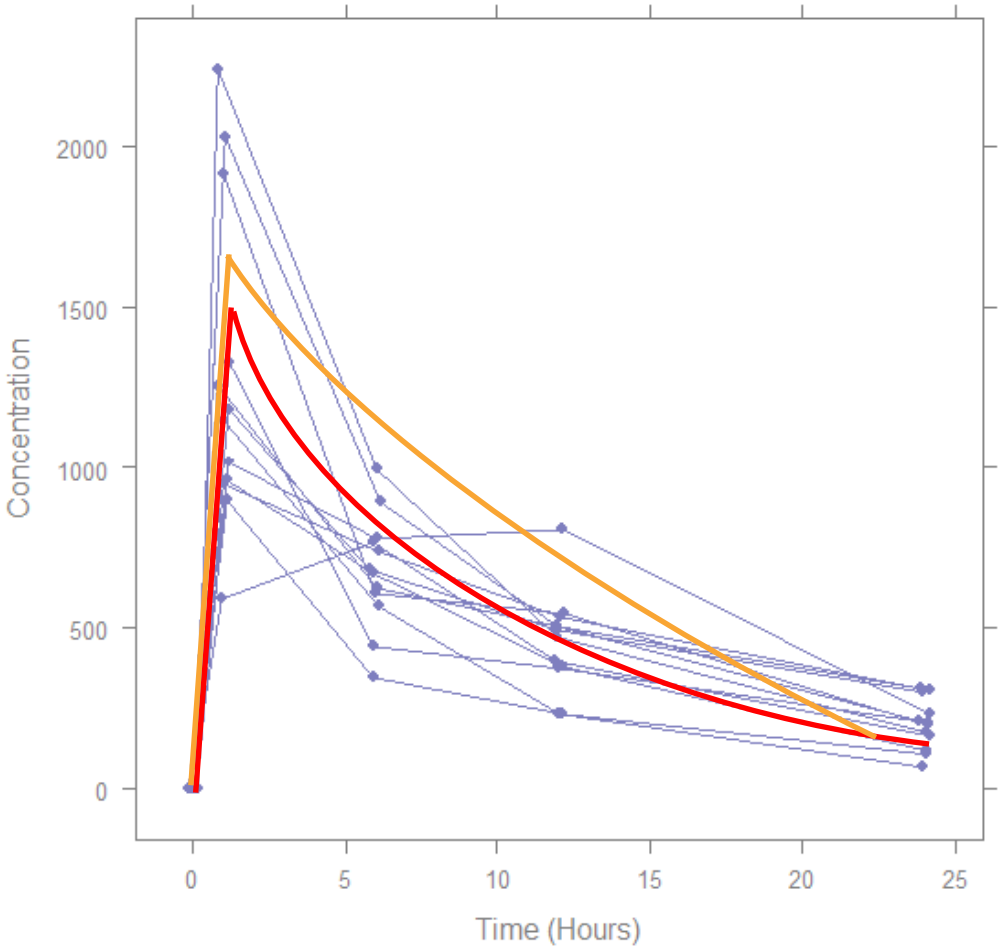
Analysis

- 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?

MANGOSOLUTIONS
data analysis that delivers



Data



Model Outputs

Models

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

ModSpace
The Project

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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 “Moogle” hints at vision!

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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”)

MANGOSOLUTIONS
data analysis that delivers

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

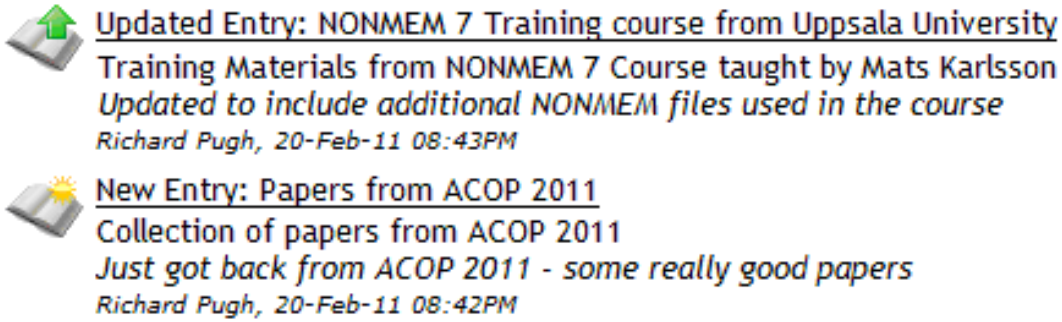
Create and Describe Entry

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



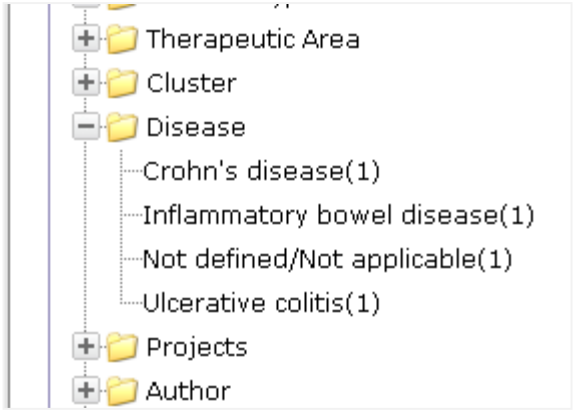
MANGOSOLUTIONS
data analysis that delivers

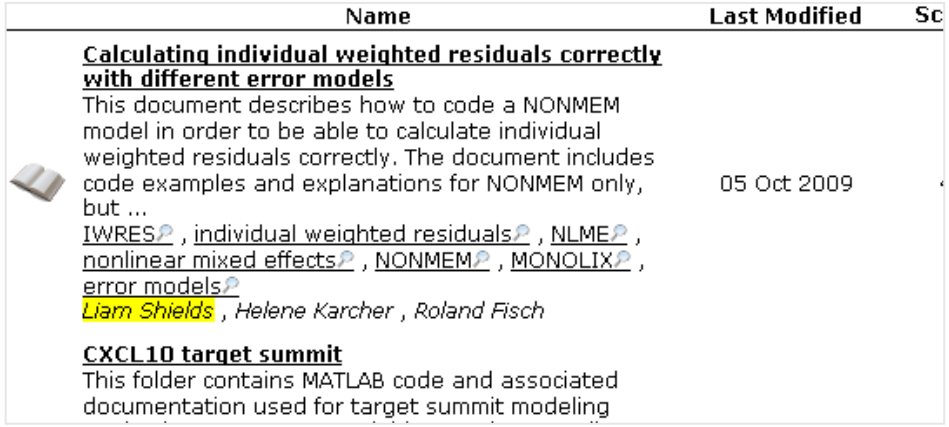
Search for Information

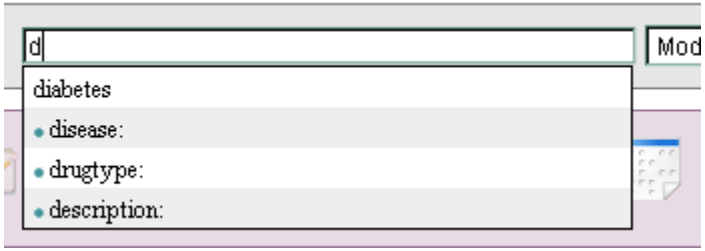
- Apache Lucene search engine behind the scenes:
  - Simple search
  - Advanced search
  - Google-Syntax search
  - Filtering of Results
  - Suggestions and Spelling-Matching

MANGOSOLUTIONS
data analysis that delivers

Search for Information







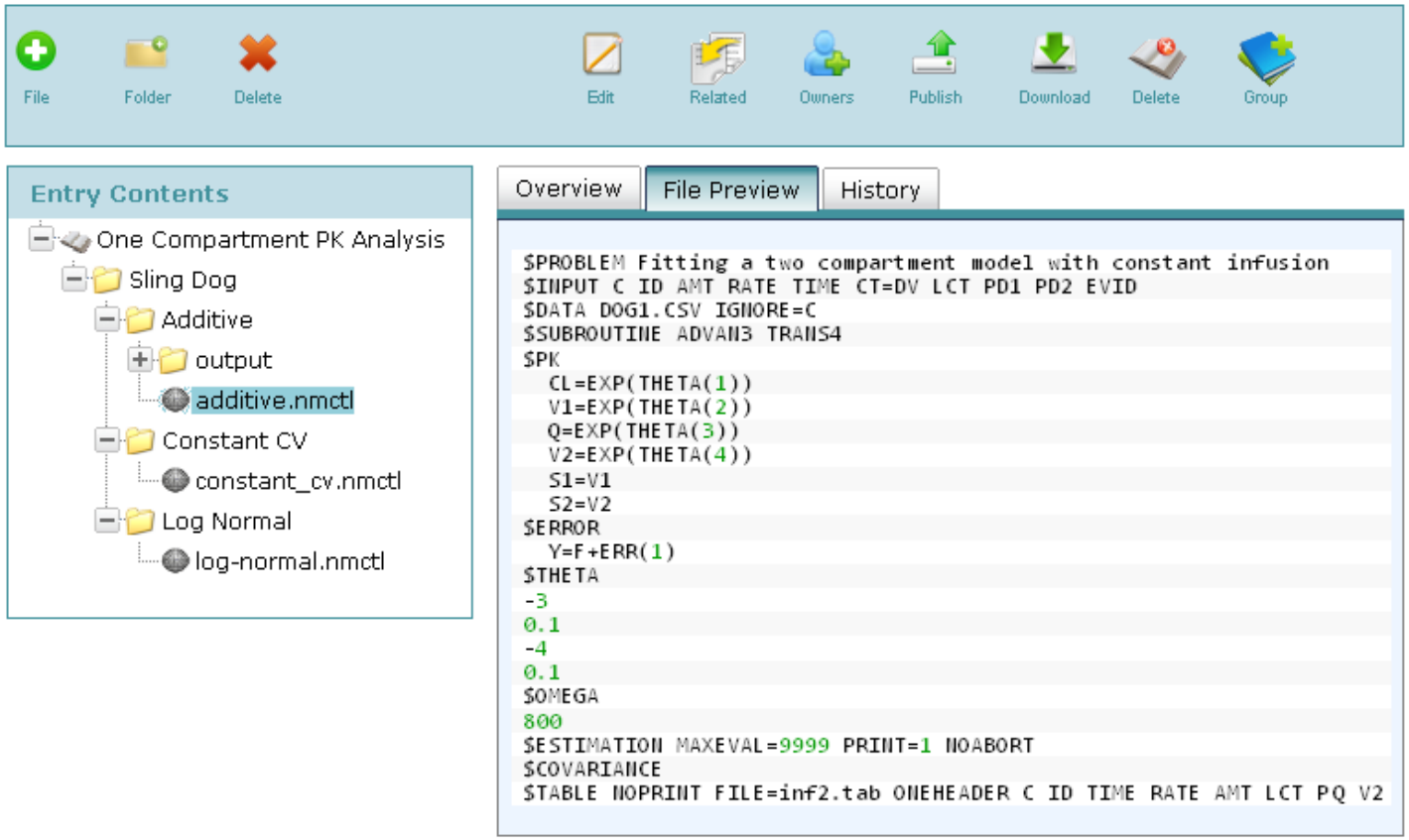
MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

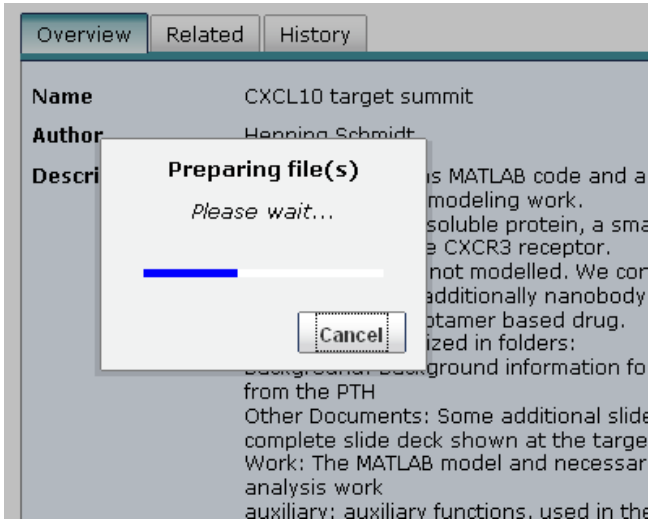
Standardised View of Information

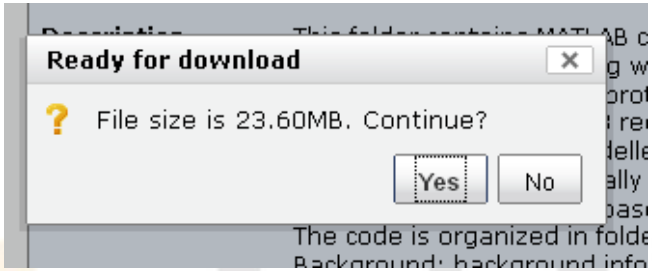


MANGOSOLUTIONS
data analysis that delivers

Download Information

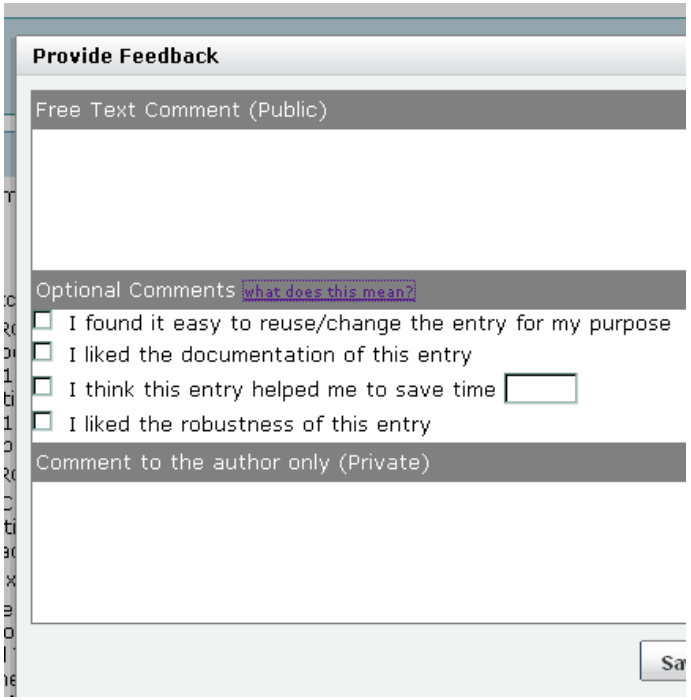
- Download single “Elements” or entire “Entry”
- Extract as Zip or work directly with version control repo
- Entry is “Bookmarked” and “Feedback” event triggered

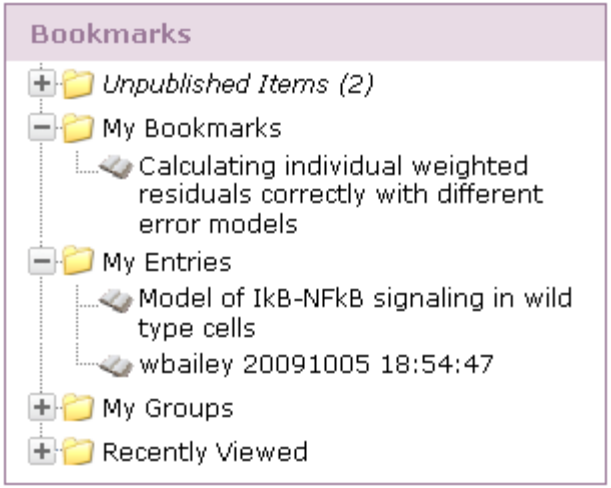




Provide Feedback

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

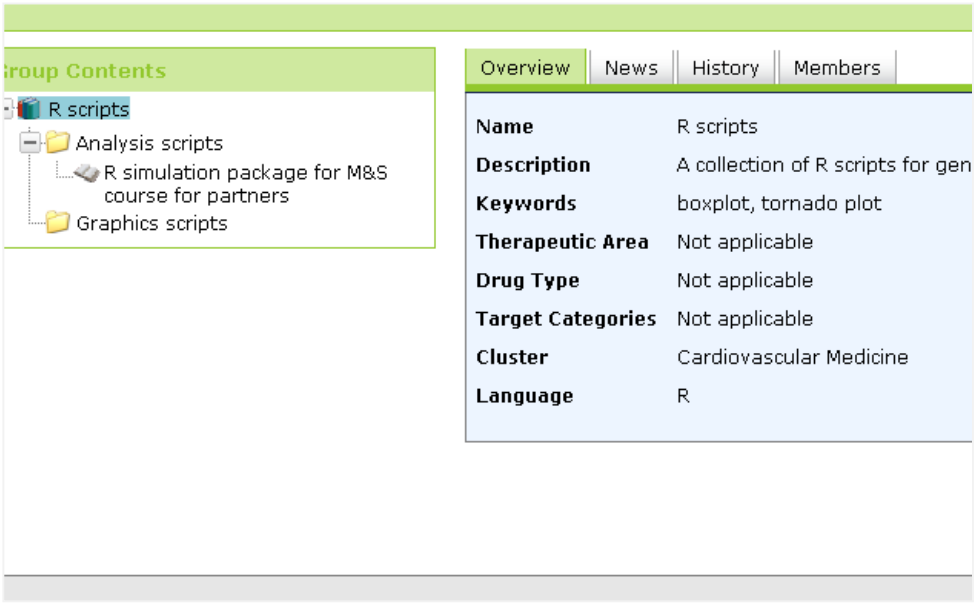




MANGOSOLUTIONS
data analysis that delivers

Create Communities

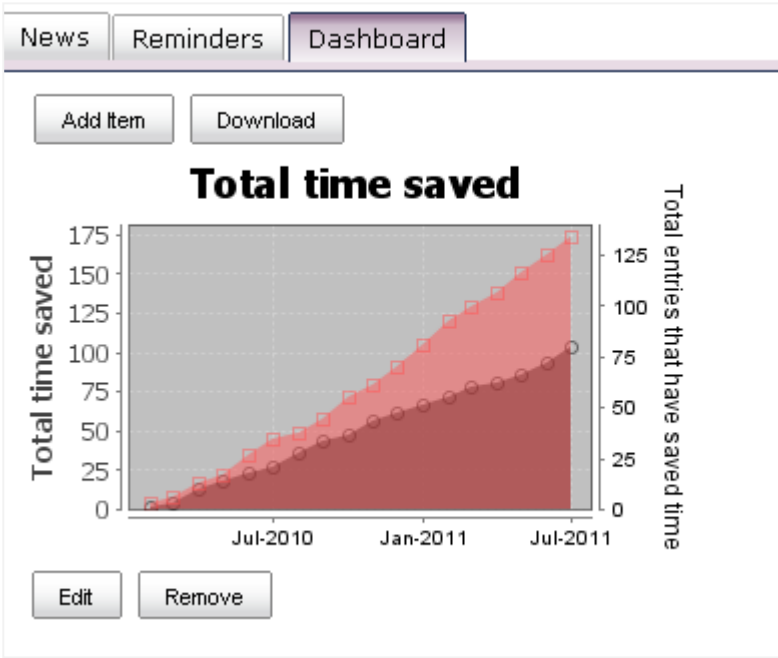
- Create “Groups”, a collection of bookmarks within a specific category
- Has it’s own membership list and metadata



MANGOSOLUTIONS
data analysis that delivers

Produce Management Reports

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



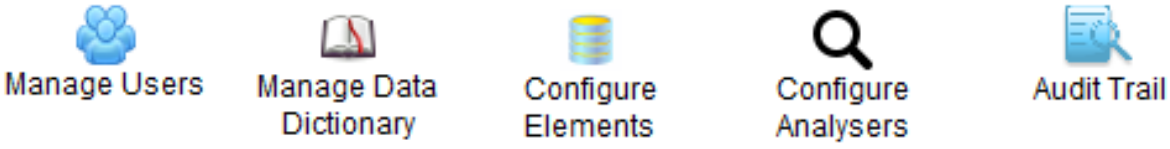
MANGOSOLUTIONS
data analysis that delivers

ModSpace
The Technical Details

MANGOSOLUTIONS
data analysis that delivers

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



MANGOSOLUTIONS
data analysis that delivers

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)

MANGOSOLUTIONS
data analysis that delivers

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”

MANGOSOLUTIONS
data analysis that delivers

ModSpace
The Alternatives

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

ModSpace
The Development Path

MANGOSOLUTIONS
data analysis that delivers

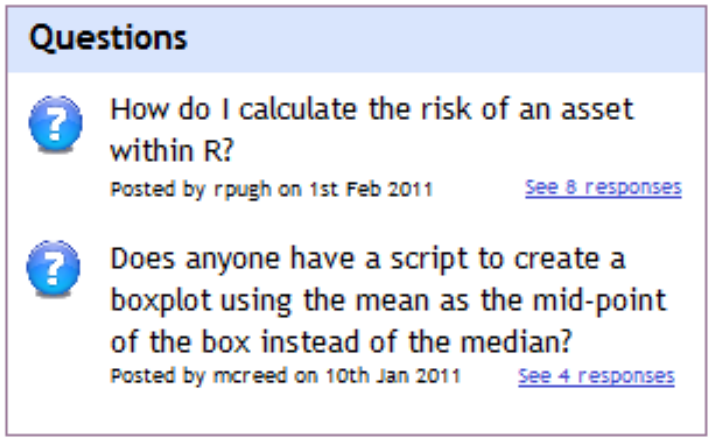
Formal Entry Structure

- 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

MANGOSOLUTIONS
data analysis that delivers

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



MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

Summary

MANGOSOLUTIONS
data analysis that delivers

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

MANGOSOLUTIONS
data analysis that delivers

Demo Modspace