HUB Application for Managing Experimental Knowledge

Girish Joglekar, Arun Giridhar, Michael McClennan, George Howlett, Gintaras V. Reklaitis

School of Chemical Engineering, ITaP
Purdue University
Outline

• Requirements of experimental data management
• Workflows
• Workflow editor
• Overview of the experimental data management system
• Summary
Experiment

• An experiment is an orderly procedure carried out with the goal of verifying, refuting, or establishing the validity of a hypothesis

• Experiments vary greatly in their scope and domain, but always rely on repeatable procedure and logical analysis of the results

• Requirements:
  - Procedures must be well documented
  - Results should be accessible
Forms of Experimental Procedure and Data

- Hand records in notebooks
- Electronically stored in unstructured format
  - Spreadsheets
  - Documents
  - Electronic lab notebooks (ELN)
- Electronically stored in structured format
  - Laboratory information management systems (LIMS) in use for only for a few years
Requirements of ‘Good Data’

• Should be structured and semantically rich
• Should be in electronic form
• Open access to facilitate sharing
  ➢ Programmatically
  ➢ Manually
• Provide data provenance to facilitate evaluation of data quality
Objective

• Objective of this work:
  ➢ Develop a workflow based system to manage experimental data
  ➢ Use case: Experiments performed on a test bed at Purdue University for drop-on-demand pharmaceutical manufacturing
Workflow

• Captures experimental procedure
  ➢ Organized as series of tasks, with each task as series of subtasks
  ➢ Resources needed to perform each task: equipment, machines, operators, utilities
  ➢ Material and information flow between subtasks

• Can be modeled graphically as a network
  ➢ Graphical models provide an intuitive, high level representation of the procedure
  ➢ Provides a logical structure to organize, store and retrieve knowledge
Sample Workflow

Raw Material

Workflow

Task

Material flow

Sink

Subtask

Data Node

Information Flow
Benefits of Workflow Based Approach

• Network can be tailored to the underlying operational procedure
  ➢ Can capture smallest level of detail

• Provides semantics to each data item

• Can be used for data entry as well as data retrieval

• An engine can be developed to
  ➢ Walk the user through each task and subtask in the specified order
  ➢ Ensure that all the information is entered
  ➢ Supervisor can sign-off on each experiment
KM System Details

- Implemented as a ‘component’ on pharmahub
- The KM system consists of:
  - **Graphical workflow editor**
    - *Tcl/Tk application*
  - **KM user interface**
    - *Server-side PHP application*
  - **Workflow execution engine**
    - *Implemented in JavaScript*
    - *Uses AJAX*
KM System Components

- Web Interface for KM
- Browser
- Workflow Builder
- Browser
- PharmaHUB
- HUBzero
- Joomla
- MySQL Server
- Web Server
- PHP
Workflow Builder

- Graphical editor to build a workflow
- Available as a tool on pharmahub
Outline

• Requirements of experimental data management

• Workflows

• Workflow editor

• Overview of the experimental data management system

• Summary
### Parameters for "Setup"...

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Units</th>
<th>Value</th>
<th>Min</th>
<th>Max</th>
<th>Group</th>
<th>In</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pump speed</td>
<td>frequency</td>
<td>rpm</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pump displacement</td>
<td>number</td>
<td>nu...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Volume strokes</td>
<td>number</td>
<td>nu...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Pump temperature</td>
<td>temperature</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Pump priming time</td>
<td>time</td>
<td>min</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
Drop-on-Demand Lab Production Facility At Purdue

Production Workflow

Calibration Workflow
Capturing Details

• Total of 30 workflow, task, subtask and raw material parameters recorded for each production run

• Metadata for three tabular data files defined in the workflow
  – *Three tabular data files (.csv) created for each experiment*
    • *Drop series data:* 13 columns
    • *NIR Raw data:* 100 columns
    • *Crystallinity data:* 3 columns

• Compressed file consisting of images of each drop in jpg format
Recording a Run
Browsing and Extracting Information

• Browsing information for a single run
  - Select a desired run
  - Workflow is displayed
  - Click on any icon, a form with entered values is displayed
  - For any data node
    ✓ Select a csv file to display as a table or download
    ✓ Select a zip file to download

• Extract information from one or more runs
  - Select parameter(s)/table column(s)
  - Select runs
  - Extract to display as a table or download
  - Draw plot
Extracting Parameters

Select Parameter(s) and Run(s) for Data Extraction

Select workflow template: TB3NozzleCalib:Experimental
Select parameters: es_3_1_Pump speed, es_
Select Experiments: 17, 18, 19, 20

Select a run

<table>
<thead>
<tr>
<th>ID</th>
<th>Purpose</th>
<th>SupervisorID</th>
<th>EmployeeID</th>
<th>LaboratoryID</th>
<th>PhaseListID</th>
<th>SpeciesListID</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Displacement 1</td>
<td>Arun Girdhar</td>
<td>Elcin Icton</td>
<td>TB3 Lab</td>
<td>Org</td>
<td>Acetone</td>
</tr>
<tr>
<td>18</td>
<td>Displacement 1</td>
<td>Arun Girdhar</td>
<td>Elcin Icton</td>
<td>TB3 Lab</td>
<td>Org</td>
<td>Acetone</td>
</tr>
<tr>
<td>19</td>
<td>Displacement 2</td>
<td>Arun Girdhar</td>
<td>Elcin Icton</td>
<td>TB3 Lab</td>
<td>Org</td>
<td>Acetone</td>
</tr>
<tr>
<td>20</td>
<td>Displacement 2</td>
<td>Arun Girdhar</td>
<td>Elcin Icton</td>
<td>TB3 Lab</td>
<td>Org</td>
<td>Acetone</td>
</tr>
<tr>
<td>30</td>
<td>Disp 2, rpm 720</td>
<td>Arun Girdhar</td>
<td>Elcin Icton</td>
<td>TB3 Lab</td>
<td>Org</td>
<td>Acetone</td>
</tr>
</tbody>
</table>

Select parameter(s) for data extraction

Task {TB3NozzleCalib,DataCapture} Parameters
- End Time

Task {TB3NozzleCalib,IVEK} Parameters
- End Time

Subtask {TB3NozzleCalib,IVEK,Setup} Parameters
- Pump speed
- Pump displacement
- Volume strokes
- Pump temperature

Save

Browse Extracted Data

Subtask,17,3,1,Pump speed,1440,rpm,
Subtask,17,3,1,Pump displacement,1.5,number,
Subtask,18,3,1,Pump speed,1680,rpm,
Subtask,18,3,1,Pump displacement,1.5,number,
Subtask,19,3,1,Pump speed,720,rpm,
Subtask,19,3,1,Pump displacement,2,number,
Subtask,20,3,1,Pump speed,720,rpm,
Subtask,20,3,1,Pump displacement,2,number,
Graphing Data

Select x and y values, and Run(s) for graphing

Select x-axis values: ec_2_2, y-axis values: ec_2_9, Experiments: 17,18,19,20

Graphing Data Presentation

- Drop number(#)
- Drop volume(µl)

Expt ID(s):
- 17
- 18
- 19
- 20

Graph showing trends in drop volume over drop number for different experiments.
Summary

• Workflow based knowledge management system is implemented on the HUBzero platform
• Graphical model of a workflow template is central to building, accessing and sharing knowledge
• Workflow based approach allows you to ‘Do it right the first time’
• Will be extended to manage scientific workflows, manufacturing recipes, business workflow