Towards Appling Social Design Principles for Organic Data Science

A task centered approach to social collaboration

Felix Michel

Prof. Dr. Yolanda Gil
University of Southern California
Information Sciences Institute
Supervisor

Prof. Dr. Florian Matthes
M.Sc. with honors Matheus Hauder
Technical University of Munich
Faculty of Informatics
Co-Supervisor

May 12, 2014

Agenda

I. **Introduction**
   I. ODS Problem Approach
   II. ODS Example

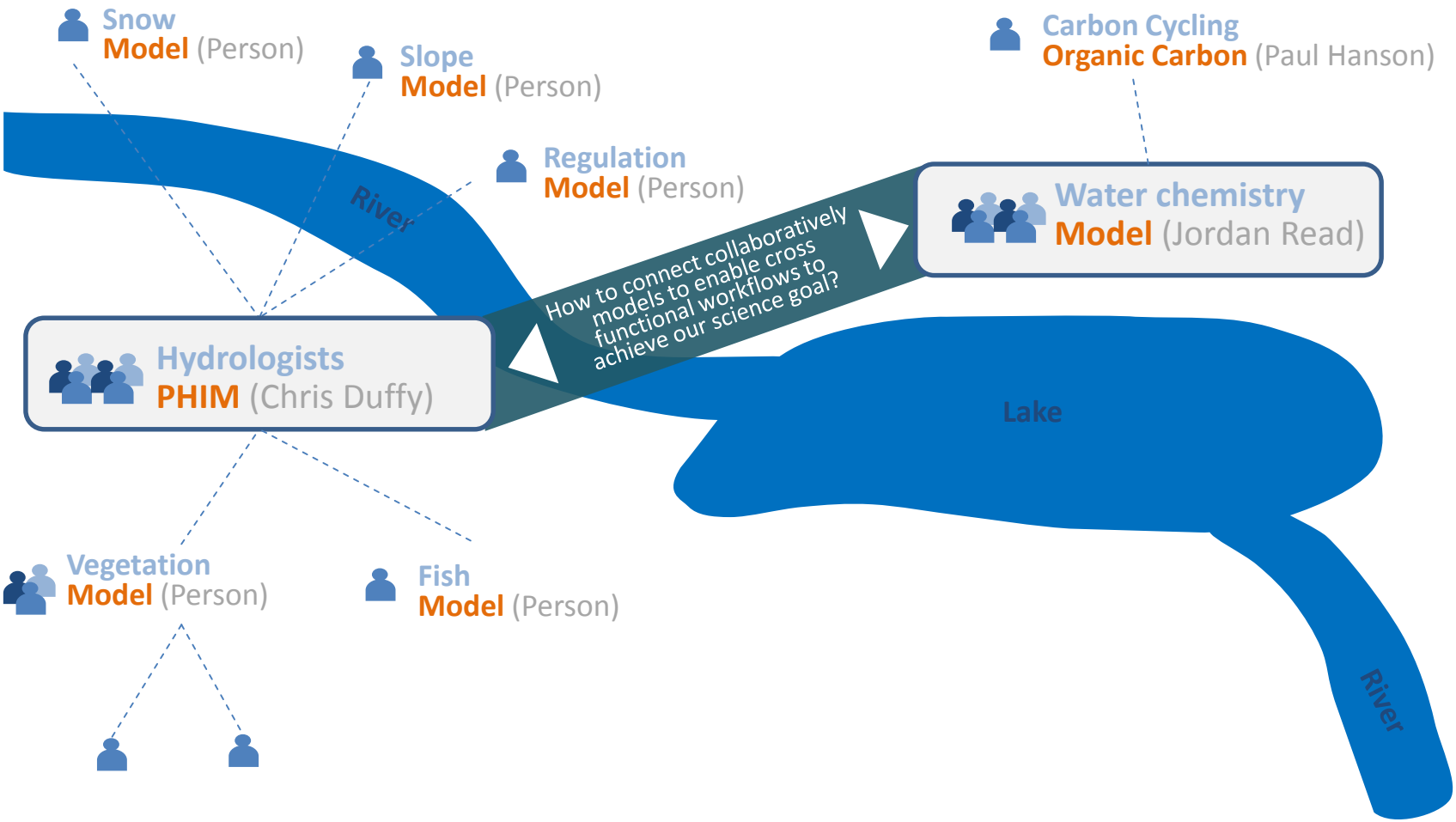
II. **Task concept**
   I. Conceptual Task Attributes
   II. Social Design Principles Addressed

III. **Next Steps**

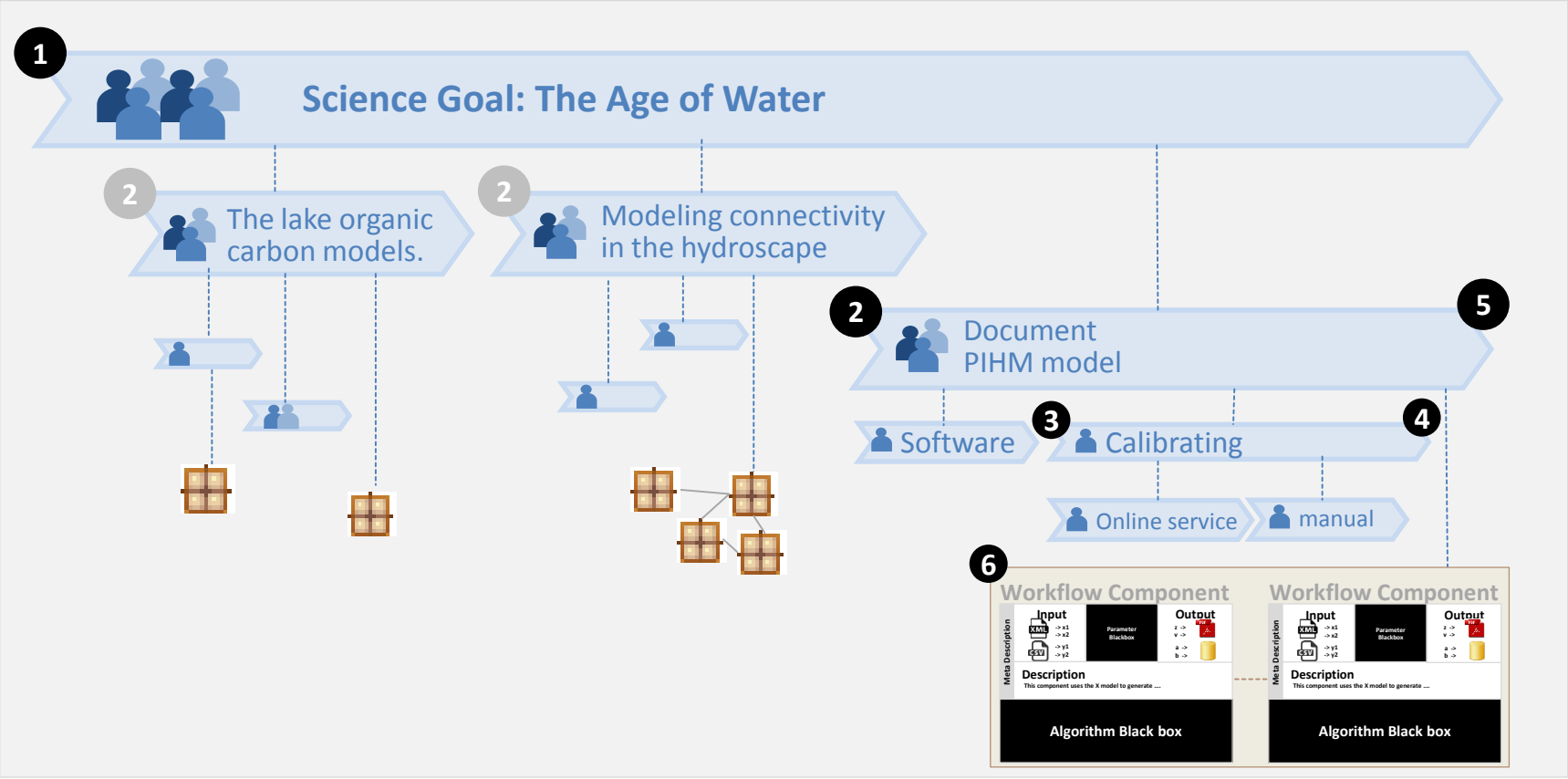
IV. **Discussion**

Organic Data Science Problem Approach

Science Goal: **The Age of Water**

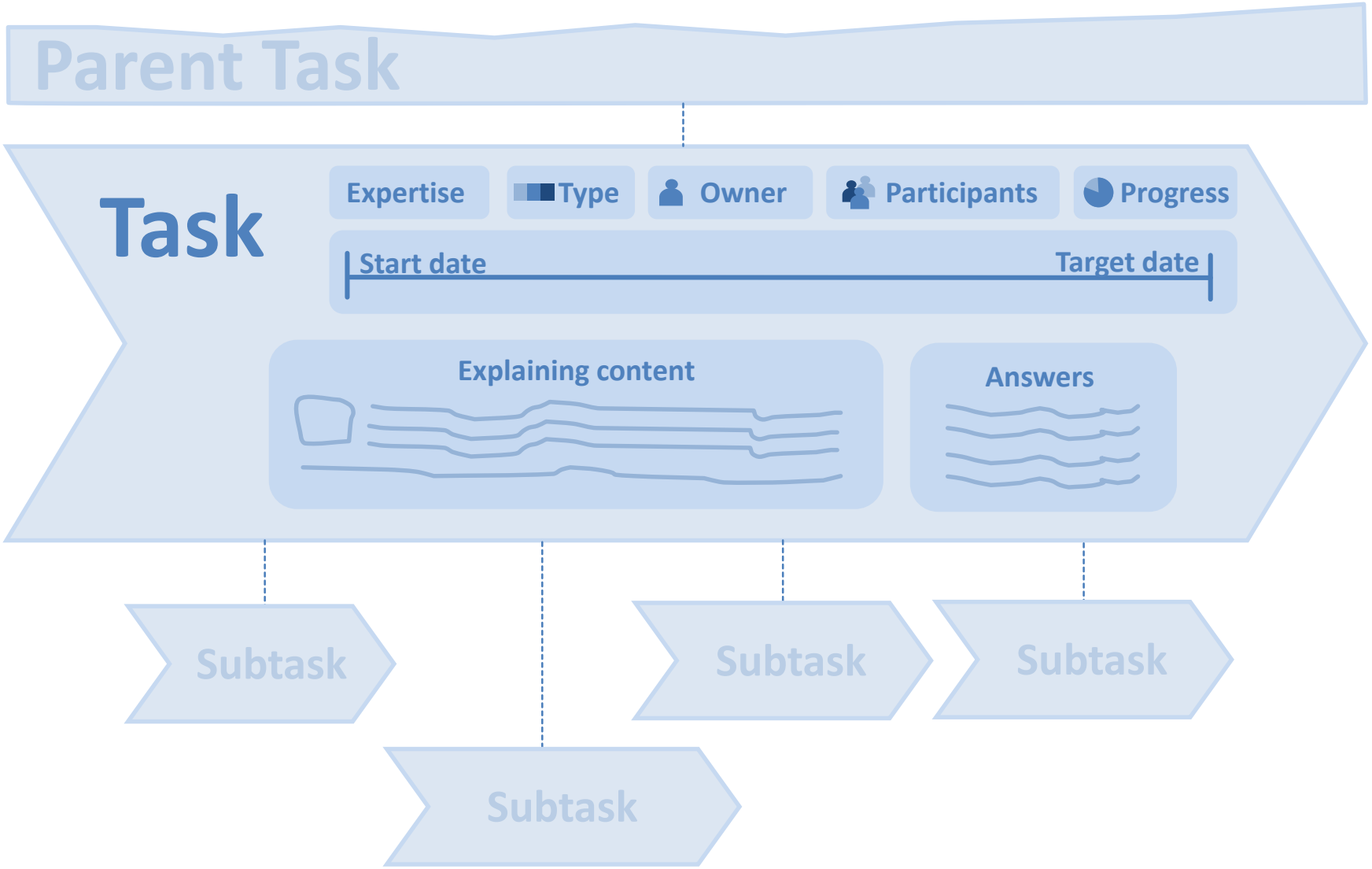


Example of Organic Data Science

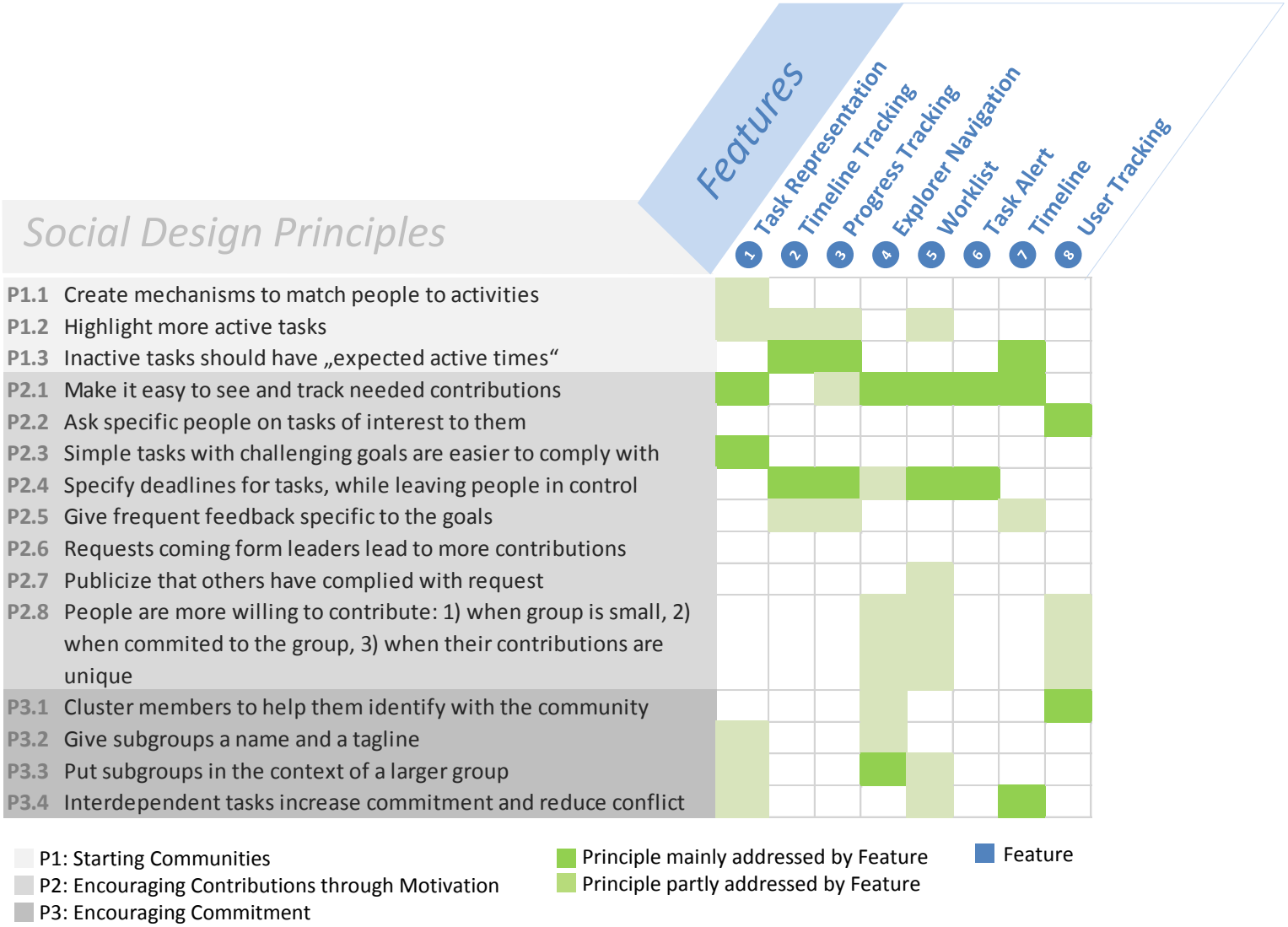


1. Defining Science Goal
2. Define Main Tasks
3. Decompose Tasks until they are solvable
4. Accomplish tasks with results
5. Aggregating Knowledge
6. Models as software component

Conceptual Task Attributes



Social Design Principles Addressed



Parent Task

Current Task

Sub-tasks

Describing Content

Structured Properties

Answers

PIHM model Documentation

Calibrating the PIHM model

- PIHM calibration as an on-line service
- PIHM manual calibration
- PIHM calibration using evolutionary

An important feature of PIHM is the utilization of geospatial data necessary to setup and run the model for a given watershed from national data sources. In simple terms, each dataset utilized in the modeling process is a geospatial map (vector or raster) with assigned properties. For example, SSURGO (USDA ref) is a national map for soil type in GIS format (see figure 1). Each soil delineated by the map is associated with a lookup table of soil texture (sand-silt-clay-bulk density-organic content) from which model hydraulic parameters are estimated. It is important to realize that the estimated parameters determined from the map, and then projected onto the numerical mesh, are considered a-priori initial guesses, and likely to have a high uncertainty. However, we have found that the most important thing about the soils data is the spatial map itself. A-priori parameters are only used as an initial guess and must be optimized as part of the model process as described below.

Structured Properties

Type	Medium-Level
Expertise	Hydrology modeling Carbon
Start Date	12/15/2014
End Date	12/15/2015
Owner	Chris
Participants	Steven Gil
Progress	70%

Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3

1

Task Representation

P2.1: Make it easy to see and track needed contributions

P2.3: Simple tasks with challenging goals are easier to comply with

PIHM model Documentation

Calibrating the PIHM model

- PIHM calibration as an on-line service
- PIHM manual calibration
- PIHM calibration using evolutionary

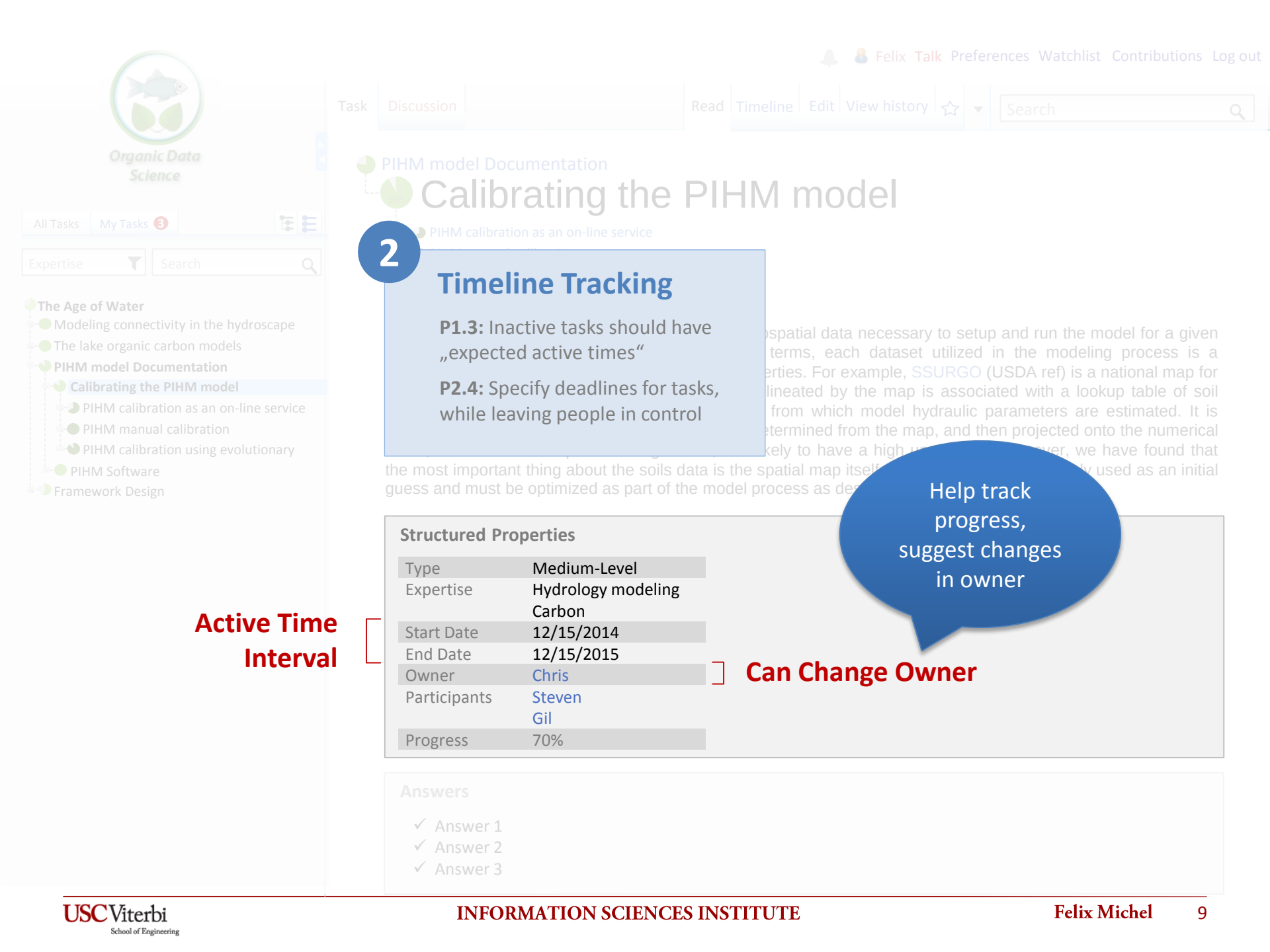
An important feature of PIHM is the utilization of geospatial data necessary to setup and run the model for a given watershed from national data sources. In simple terms, each dataset utilized in the modeling process is a geospatial map (vector or raster) with assigned properties. For example, SSURGO (USDA ref) is a national map for soil type in GIS format (see figure 1). Each soil delineated by the map is associated with a lookup table of soil texture (sand-silt-clay-bulk density-organic content) from which model hydraulic parameters are estimated. It is important to realize that the estimated parameters determined from the map, and then projected onto the numerical mesh, are considered a-priori initial guesses, and likely to have a high uncertainty. However, we have found that the most important thing about the soils data is the spatial map itself. A-priori parameters are only used as an initial guess and must be optimized as part of the model process as described below.

Structured Properties

Type	Medium-Level
Expertise	Hydrology modeling Carbon
Start Date	12/15/2014
End Date	12/15/2015
Owner	Chris
Participants	Steven Gil
Progress	70%

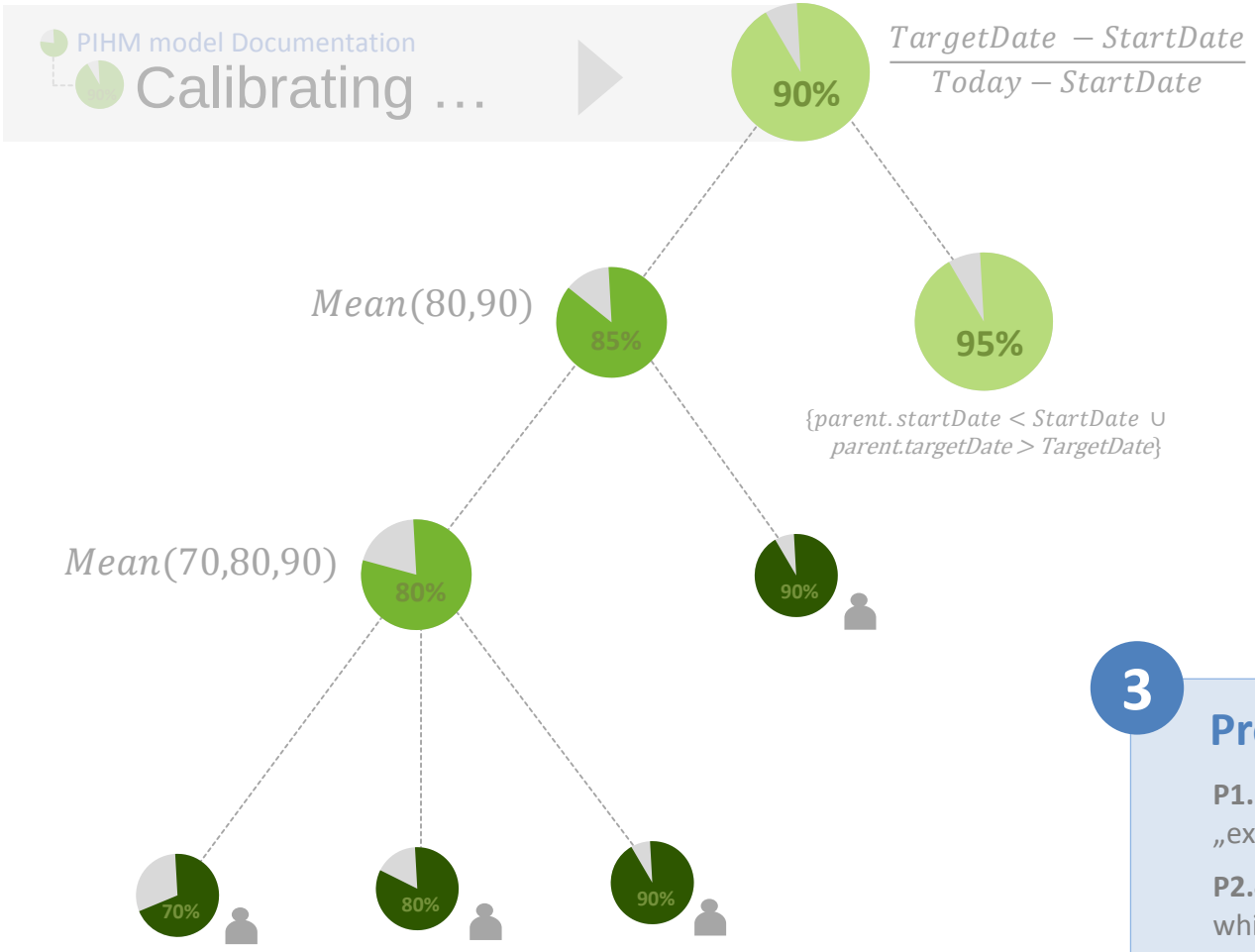
Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3


Organic Data Science
Felix Talk Preferences Watchlist Contributions Log out
Task Discussion Read Timeline Edit View history Search
All Tasks My Tasks 3
Expertise Search
The Age of Water
Modeling connectivity in the hydroscape
The lake organic carbon models
PIHM model Documentation
Calibrating the PIHM model
PIHM calibration as an on-line service
PIHM manual calibration
PIHM calibration using evolutionary
PIHM Software
Framework Design
PIHM model Documentation
Calibrating the PIHM model
PIHM calibration as an on-line service
2
Timeline Tracking
P1.3: Inactive tasks should have „expected active times“
P2.4: Specify deadlines for tasks, while leaving people in control
Structured Properties
Type Medium-Level
Expertise Hydrology modeling Carbon
Start Date 12/15/2014
End Date 12/15/2015
Owner Chris
Participants Steven Gil
Progress 70%
Active Time Interval
Can Change Owner
Help track progress, suggest changes in owner
Answers
Answer 1
Answer 2
Answer 3

Task progress measurement concept

Different measurement levels are used to ensure a simple and sufficiently accurate measurement.



Rule: The abstraction level of a parent node/task can be equal or higher but not lower.

90% **High-level Task**
Low accuracy
Estimated by Start/End-date

90% **Medium-level Task**
Medium accuracy
Aggregated Estimation

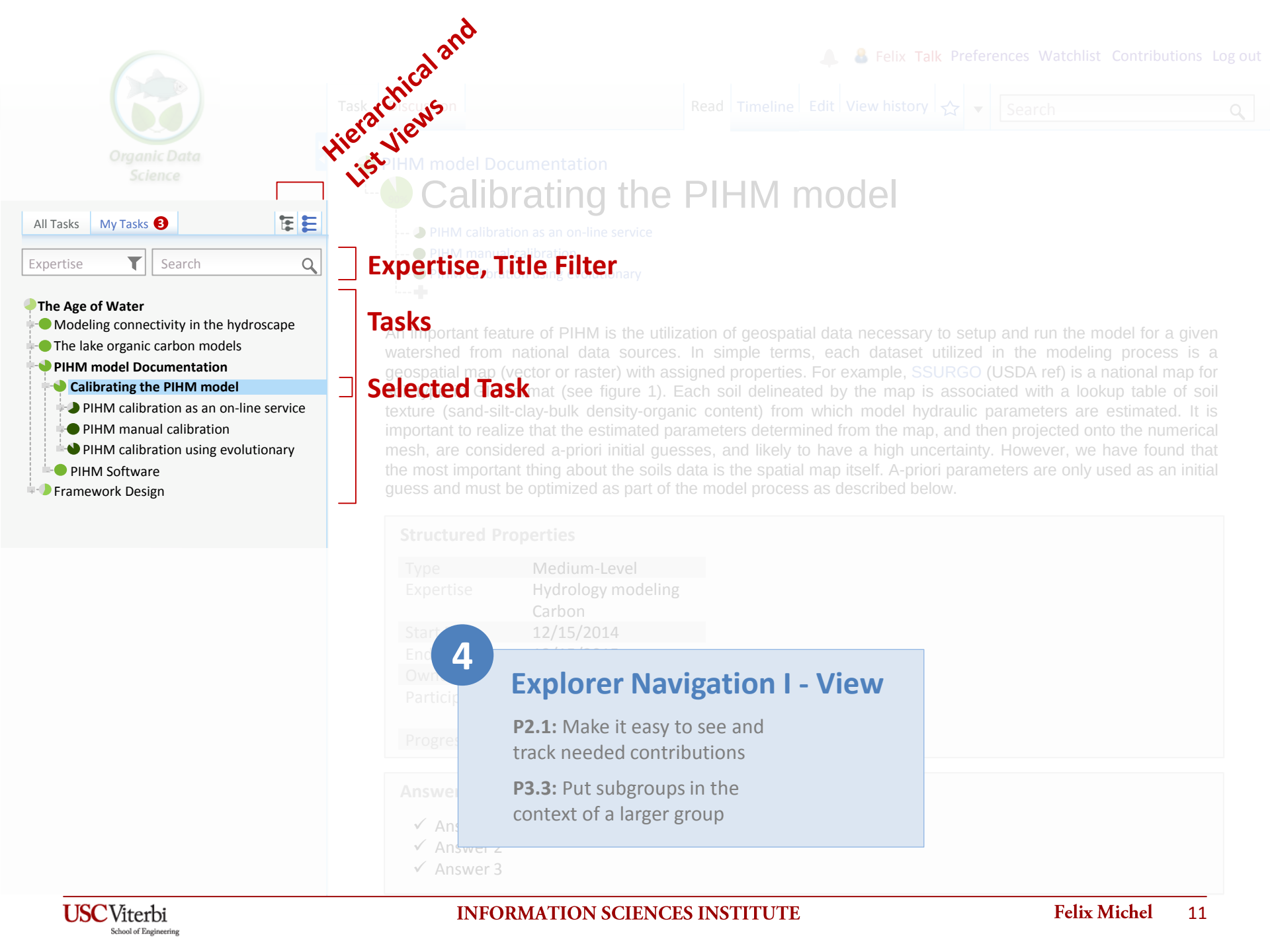
90% **Low-level Task**
High accuracy
Estimated by users

3

Progress Tracking

P1.3: Inactive tasks should have „expected active times“

P2.4: Specify deadlines for tasks, while leaving people in control

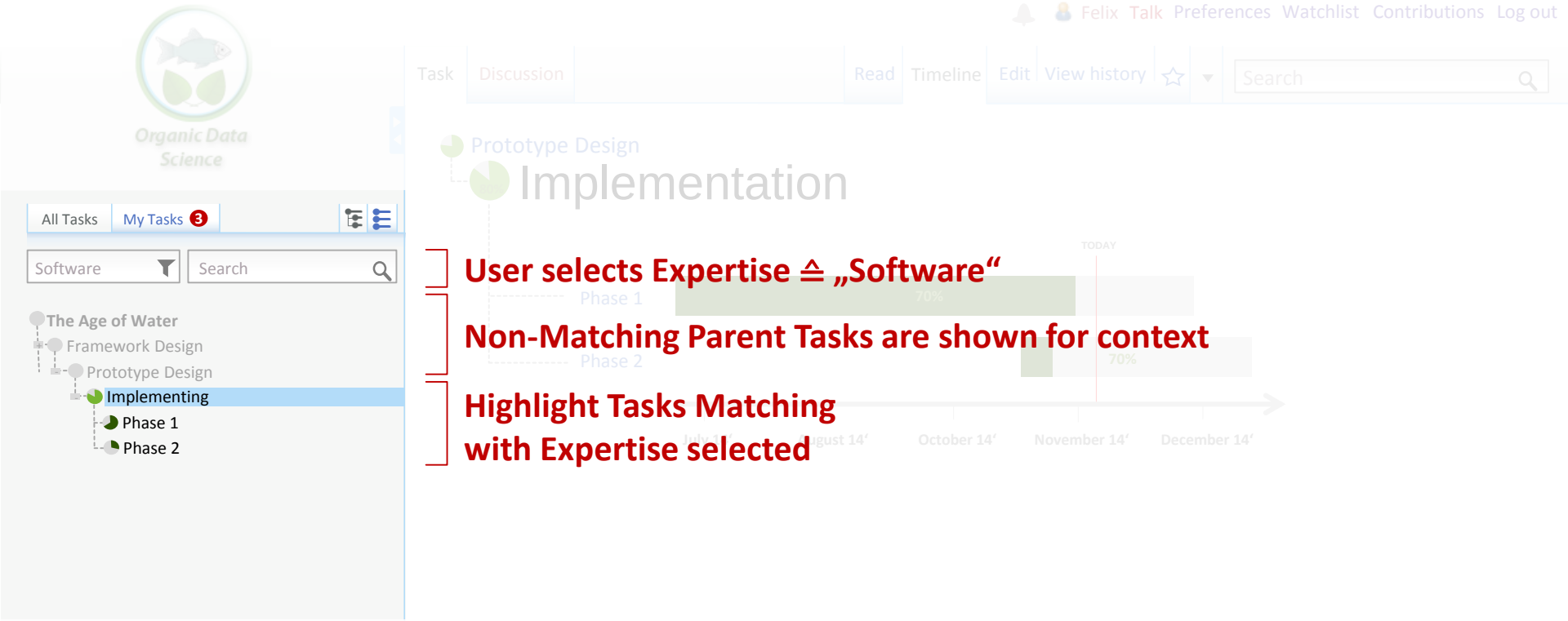


4

Explorer Navigation I - View

P2.1: Make it easy to see and track needed contributions

P3.3: Put subgroups in the context of a larger group



4

Explorer Navigation II - Filter

P2.1: Make it easy to see and track needed contributions

All Tasks

Worklist

Hierarchical View

All Tasks | My Tasks 3

Expertise | Search

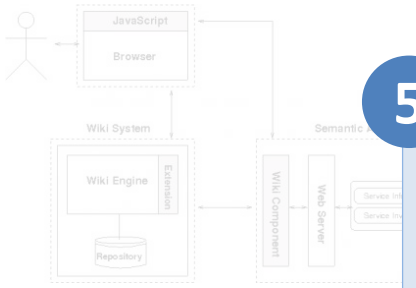
- The Age of Water
  - Framework Design
  - Prototype Design
  - Implementing
    - Phase 1
    - Phase 2

Prototype Design

Implementation

Phase 1

Phase 2



5

Worklist I – Hierarchical View

P2.1: Make it easy to see and track needed contributions

P2.4: Specify deadlines for tasks, while leaving people in control

Structured Properties

Task Type	Medium-Level
Field	Software Development
Start Date	12/15/2014
End Date	12/15/2015
Owner	Felix
Participants	Gil Varun
Progress	80%

Task Owner

Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3

Organic Data Science

Felix Talk Preferences Watchlist Contributions Log out

Task Discussion Read Timeline Edit View history Search

List View

All Tasks | My Tasks 3

Expertise | Search

Sorted By

Implementing
5 Month left

Phase 1
2 Month left

Phase 2
4 Month left

Prototype Design
Implementation
Phase 1
Phase 2

Sort Category and Option

Sortable Task List

5

Worklist II – List View

P2.1: Make it easy to see and track needed contributions

P2.4: Specify deadlines for tasks, while leaving people in control

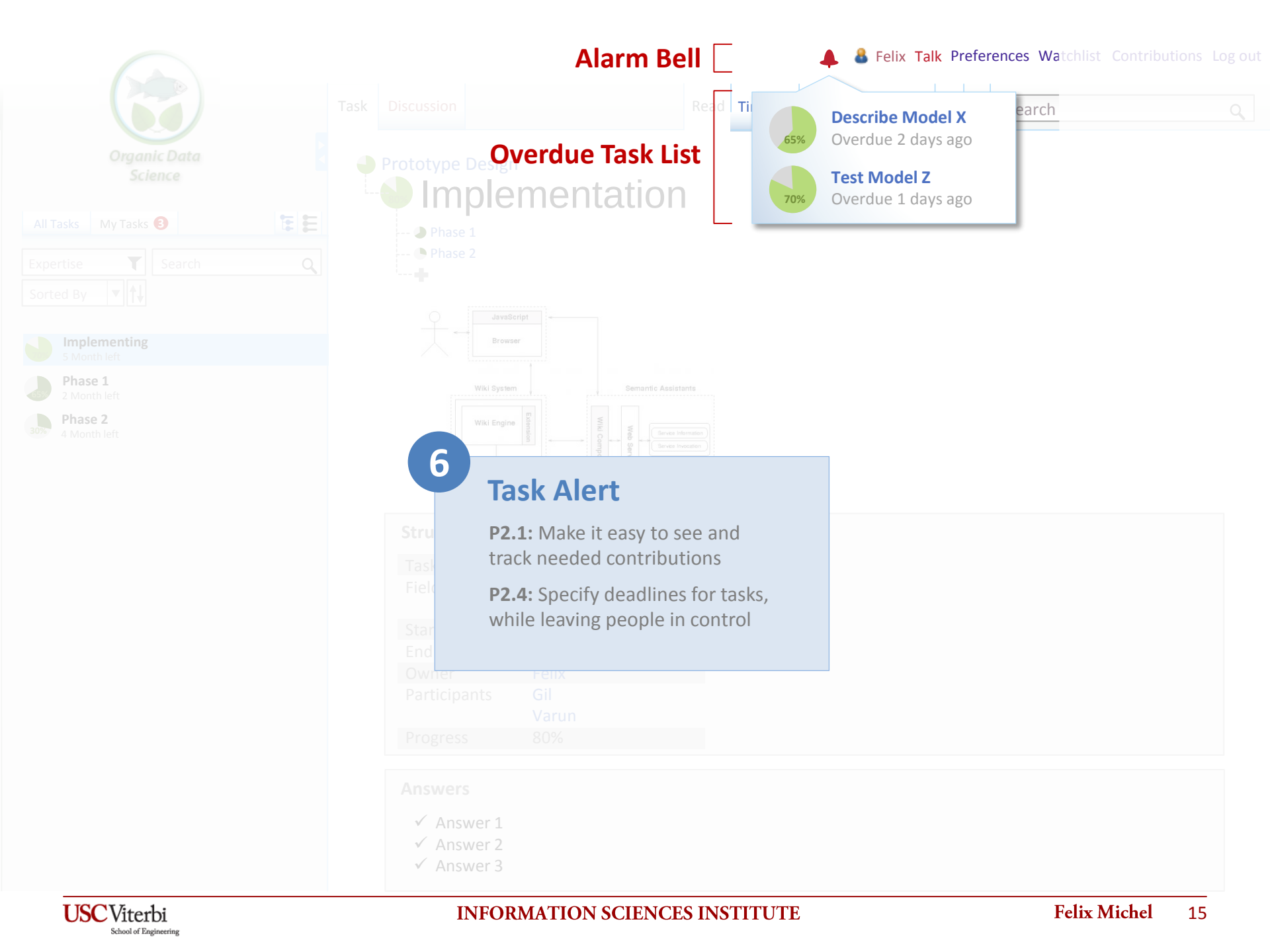
Structured Properties

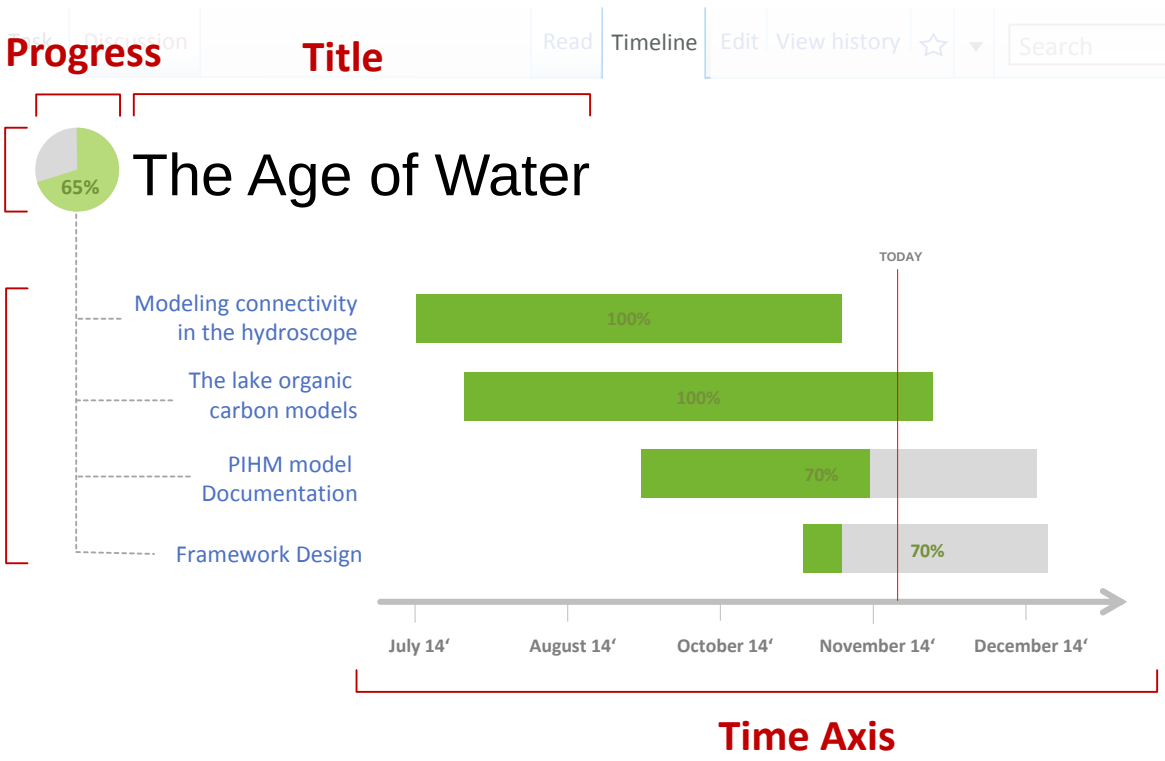
Task Type	Medium-Level
Field	Software Development
Start Date	12/15/2014
End Date	12/15/2015
Owner	Felix
Participants	Gil Varun
Progress	80%

Task Owner

Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3


Alarm Bell
Overdue Task List
Describe Model X
Overdue 2 days ago
65%
Test Model Z
Overdue 1 days ago
70%
6
Task Alert
P2.1: Make it easy to see and track needed contributions
P2.4: Specify deadlines for tasks, while leaving people in control



7

Timeline

P1.3: Inactive tasks should have „expected active times“

P2.1: Make it easy to see and track needed contributions

P2.4: Specify deadlines for tasks, while leaving people in control

Username

User picture and expertise

Active Tasks

Future Tasks

Task History

Paul Hanson

Carbon Cycling | Expertise2 | Expertise3 | Expertise4 | Expertise5

Current Tasks 4

80% **High-level task name**
Some tasks description based on the task describing content

90% **Medium-level task name**
Some tasks description based on the task describing content

60% **Low-level task name**
Some tasks description based on the task describing content

Future Tasks 1

0% **Medium-level task name**
Some tasks description based on the task describing content

Completed Tasks 1

100% **High-level task name**
Some tasks description based on the task describing content

8

User Tracking I

P2.2: Ask specific people on tasks of interest to them

P3.1: Cluster members to help them identify with the community

Organic Data Science

All Tasks | My Tasks 3

Expertise | Search

- The Age of Water
  - Modeling connectivity in the hydroscape
  - The lake organic carbon models
  - PIHM model Documentation
    - Calibrating the PIHM model
      - PIHM calibration as an on-line service
      - PIHM manual calibration
      - PIHM calibration using evolutionary
    - PIHM Software
  - Framework Design

Felix Talk Preferences Watchlist Contributions Log out

User | Discussion | Read | Timeline | Edit | View history | Search

Paul Hanson

Carbon Cycling | Expertise2 | Expertise3 | Expertise4 | Expertise5

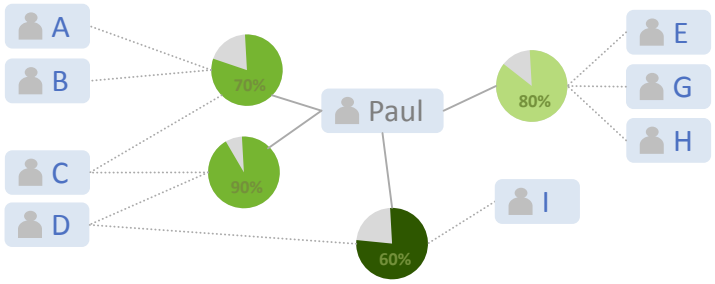
Completed Tasks 1

High-level task name

Some tasks description based on the task describing content

Task-Person Visualization

Paul's Task Network



8

User Tracking II

P2.2: Ask specific people on tasks of interest to them

P3.1: Cluster members to help them identify with the community

Next Steps

- ✓ Discuss presented concept

  Organic Data Science wiki -> Framework Design

- ✓ Iteratively implement presented concept

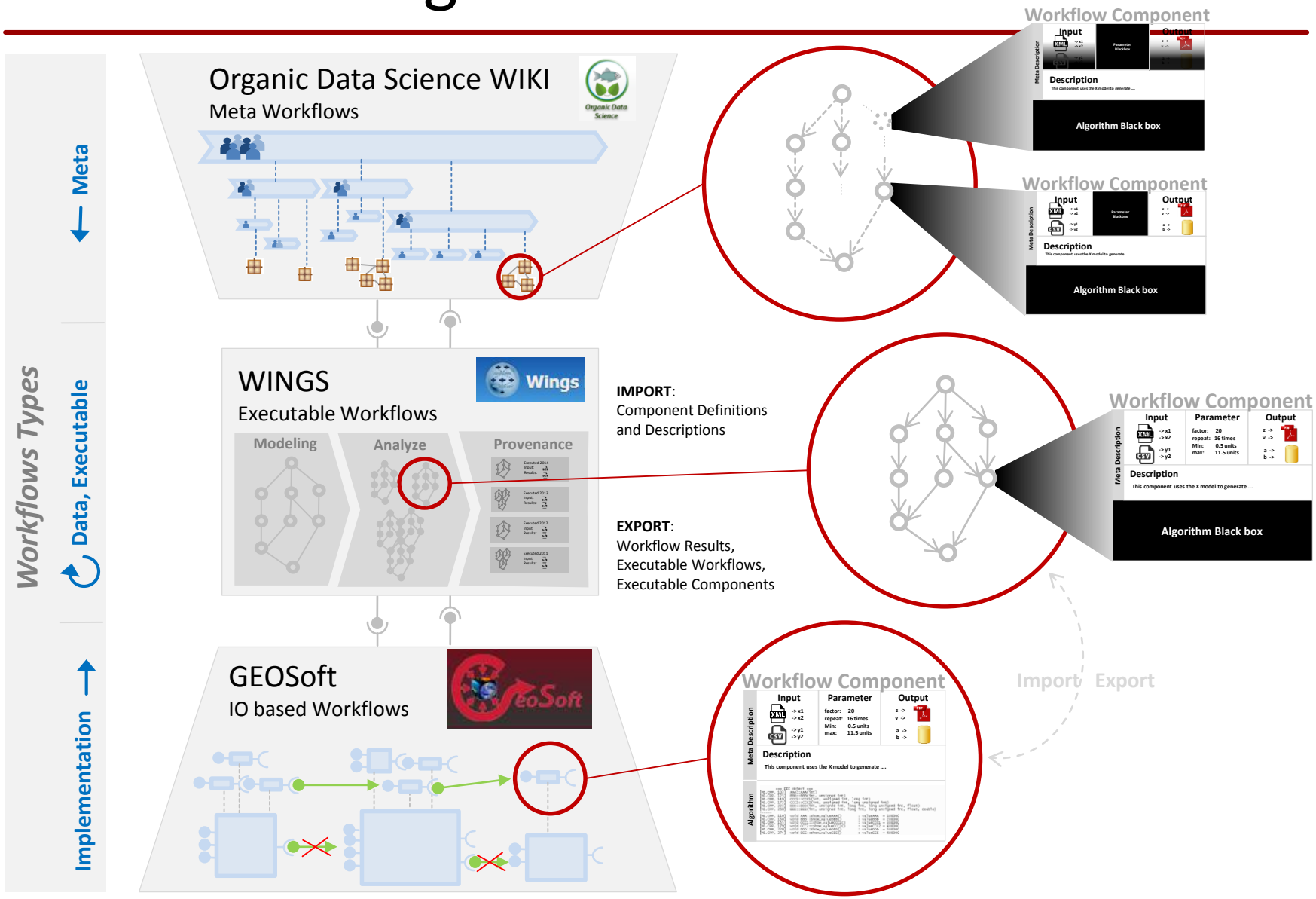
  1) Task Representation, 2) Explorer Navigation, 3) Progress Tracking, 4) Timeline, …

- ✓ Evaluate Implementation

Backup

- ✓ ODS Context
- ✓ Mockups without Annotations
- ✓ Component Mockup

Context of Organic Data Science Wiki



PIHM model Documentation

Calibrating the PIHM model

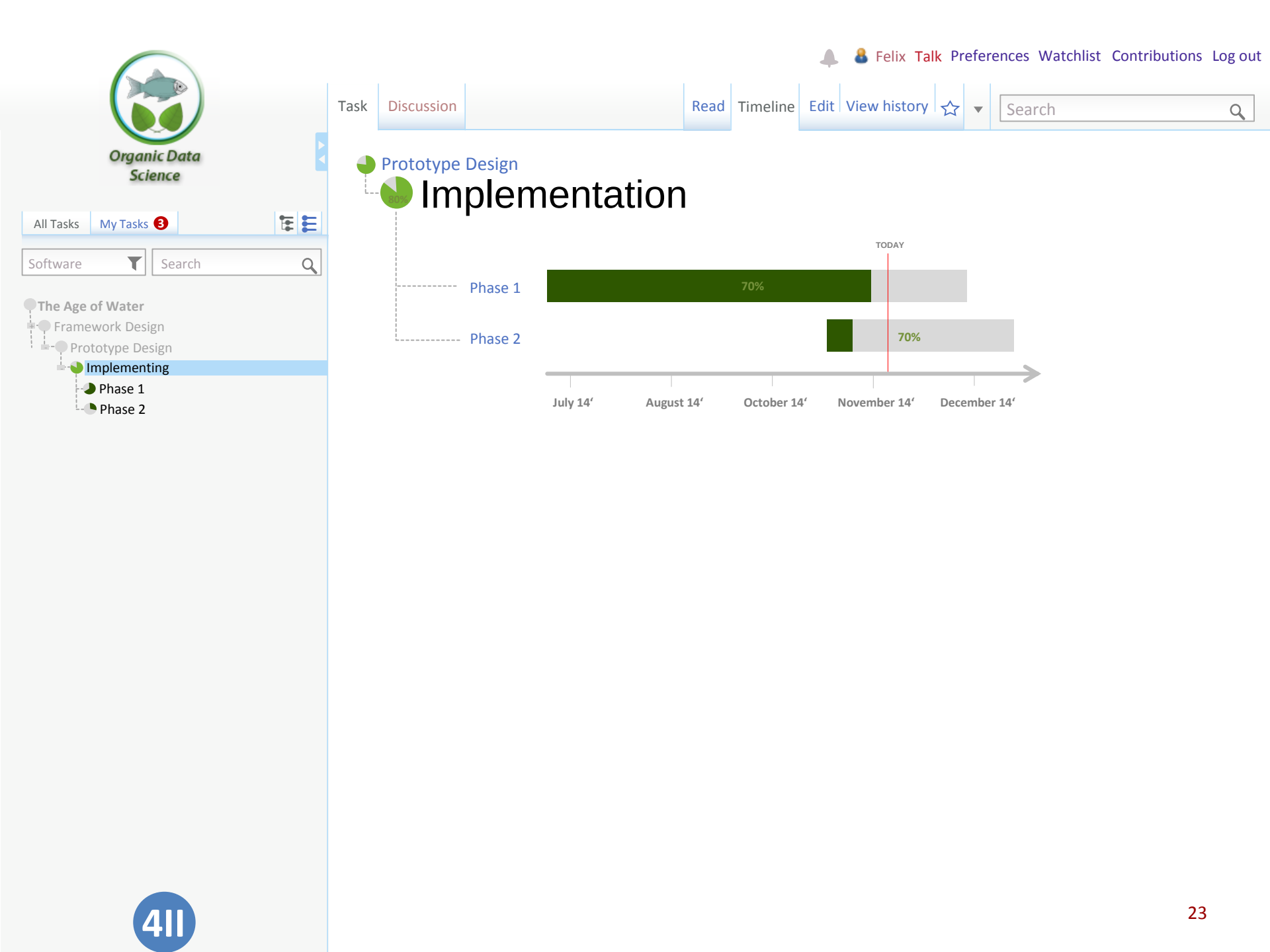
An important feature of PIHM is the utilization of geospatial data necessary to setup and run the model for a given watershed from national data sources. In simple terms, each dataset utilized in the modeling process is a geospatial map (vector or raster) with assigned properties. For example, SSURGO (USDA ref) is a national map for soil type in GIS format (see figure 1). Each soil delineated by the map is associated with a lookup table of soil texture (sand-silt-clay-bulk density-organic content) from which model hydraulic parameters are estimated. It is important to realize that the estimated parameters determined from the map, and then projected onto the numerical mesh, are considered a-priori initial guesses, and likely to have a high uncertainty. However, we have found that the most important thing about the soils data is the spatial map itself. A-priori parameters are only used as an initial guess and must be optimized as part of the model process as described below.

Structured Properties

Type	Medium-Level
Expertise	Hydrology modeling Carbon
Start Date	12/15/2014
End Date	12/15/2015
Owner	Chris
Participants	Steven Gil
Progress	70%

Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3


Felix Talk Preferences Watchlist Contributions Log out
Organic Data Science
Task Discussion Read Timeline Edit View history Search
All Tasks My Tasks 3
Software Search
The Age of Water
Framework Design
Prototype Design
Implementing
Phase 1
Phase 2
Prototype Design
Implementation
80%
TODAY
Phase 1
70%
Phase 2
70%
July 14‘
August 14‘
October 14‘
November 14‘
December 14‘
411
23

Organic Data Science

Felix Talk Preferences Watchlist Contributions Log out

Task Discussion Read Timeline Edit View history Search

All Tasks My Tasks 3

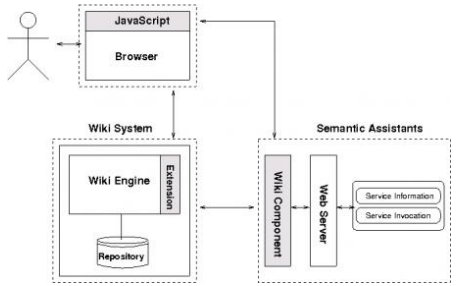
Expertise Search

- The Age of Water
  - Framework Design
    - Prototype Design
      - Implementing
        - Phase 1
        - Phase 2

Prototype Design

Implementation

- Phase 1
- Phase 2



Structured Properties

Task Type	Medium-Level
Field	Software Development
Start Date	12/15/2014
End Date	12/15/2015
Owner	Felix
Participants	Gil Varun
Progress	80%

Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3

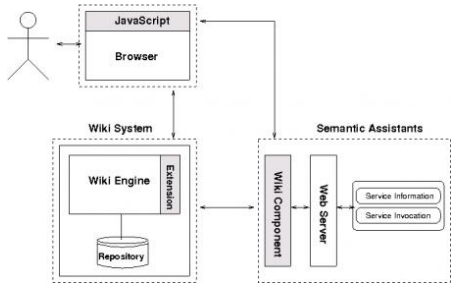
51

- **Implementing** 5 Month left
- **Phase 1** 2 Month left
- **Phase 2** 4 Month left

Prototype Design

Implementation

- Phase 1
- Phase 2



Structured Properties

Task Type	Medium-Level
Field	Software Development
Start Date	12/15/2014
End Date	12/15/2015
Owner	Felix
Participants	Gil Varun
Progress	80%

Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3

Organic Data Science

Felix Talk Preferences Watchlist Contributions Log out

Task Discussion Read Ti earch

Describe Model X
Overdue 2 days ago

Test Model Z
Overdue 1 days ago

All Tasks My Tasks 3

Expertise Search

Sorted By

Implementing
5 Month left

Phase 1
2 Month left

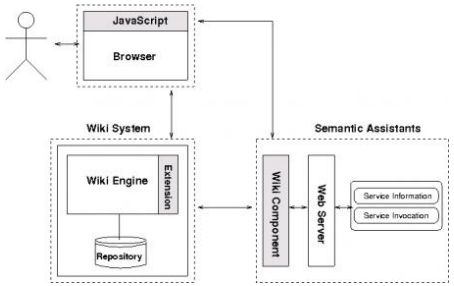
Phase 2
4 Month left

Prototype Design

Implementation

Phase 1

Phase 2



Structured Properties

Task Type	Medium-Level
Field	Software Development
Start Date	12/15/2014
End Date	12/15/2015
Owner	Felix
Participants	Gil Varun
Progress	80%

Answers

- ✓ Answer 1
- ✓ Answer 2
- ✓ Answer 3

6

Organic Data Science

Felix Talk Preferences Watchlist Contributions Log out

Task | Discussion | Read | Timeline | Edit | View history | Search

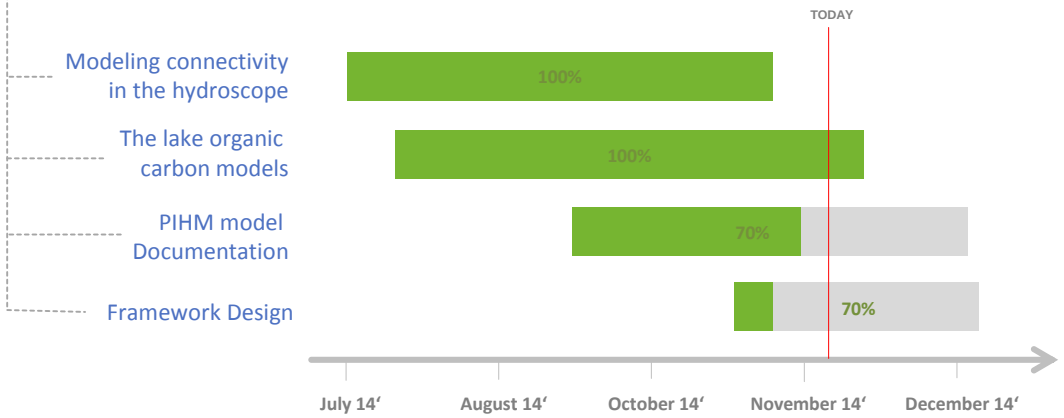
All Tasks | My Tasks 3

Expertise | Search

- The Age of Water
  - Modeling connectivity in the hydroscape
  - The lake organic carbon models
  - PIHM model Documentation
    - Calibrating the PIHM model
      - PIHM calibration as an on-line service
      - PIHM manual calibration
      - PIHM calibration using evolutionary
    - PIHM Software
  - Framework Design

The Age of Water

65%



7

Organic Data Science

Felix Talk Preferences Watchlist Contributions Log out

User | Discussion | Read | Timeline | Edit | View history | Search

All Tasks | My Tasks 3

Expertise | Search

- The Age of Water
  - Modeling connectivity in the hydroscape
  - The lake organic carbon models
  - PIHM model Documentation
    - Calibrating the PIHM model
      - PIHM calibration as an on-line service
      - PIHM manual calibration
      - PIHM calibration using evolutionary
    - PIHM Software
  - Framework Design

Paul Hanson

Carbon Cycling | Expertise2 | Expertise3 | Expertise4 | Expertise5

Current Tasks 4

80% **High-level task name**
Some tasks description based on the task describing content

90% **Medium-level task name**
Some tasks description based on the task describing content

70% **Medium-level task name**
Some tasks description based on the task describing content

60% **Low-level task name**
Some tasks description based on the task describing content

Future Tasks 1

0% **Medium-level task name**
Some tasks description based on the task describing content

Completed Tasks 1

100% **High-level task name**
ome tasks description based on the task describing content

81

Organic Data Science

Felix Talk Preferences Watchlist Contributions Log out

User Discussion Read Timeline Edit View history Search

All Tasks My Tasks 3

Expertise Search

- The Age of Water
  - Modeling connectivity in the hydroscape
  - The lake organic carbon models
  - PIHM model Documentation
    - Calibrating the PIHM model
      - PIHM calibration as an on-line service
      - PIHM manual calibration
      - PIHM calibration using evolutionary
    - PIHM Software
  - Framework Design

Paul Hanson

Carbon Cycling Expertise2 Expertise3 Expertise4 Expertise5

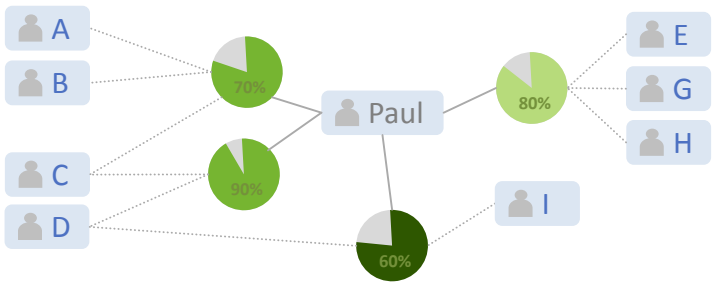
Completed Tasks 1

100%

High-level task name

Some tasks description based on the task describing content

Paul's Task Network



811

PIHM model Documentation

PIHM Software Component

Conceptual Input Variables

Name	Type	File
x1	float	TextToAnalyze
x2	float	TextToAnalyze
y1	integer	germanWords

Input files

Name	Type
TextToAnalyze	PDF
germanWords	CSV

Conceptual Output Variables

Name	Type	File
z1	integer	restultset
z2	float	restultset
z3	integer	restultset

Output files

Name	Type
report	PDF
resultset	CSV