Data Import Specification

## Task

Take an unknown formatted text file containing specific data and map it correctly into a database table.

## Assumptions

Precipitation is rounded to the nearest integer in mm

Lines are terminated with CRLF

Description starts at line X and finishes at line Y

Headers start at line X and finish at line Y

Data is taken on the same date (1st of every month)

Data is always separated by a space

## Technical Information

Phase 1

import the data into a simple table

[Xref]

[Yref]

[Date]

[Value]

Phase 2

**Unit Testing (this would be done in conjunction with developing methods)**

Better error reporting when uploading

Separate processing library out into a Business Logic Layer

Separate models out into a Data Access Layer

Validation on input fields for comma separated values.

Allow ranges for data.

Async upload to prevent timeout on large data sets, and to provide progress.

Import other information contained within the file into other tables and normalize the tables for data drill downs

[DataSet]

[DataSetID]

[DataSetDescription]

[DataSetDetails]

[DataSetDetailsID]

[DataSetID]

[LongtitudeX]

[LongtitudeY]

[LatitudeX]

[LatitudeY]

[GridX]

[GridY]

[Boxes]

[YearRange]

[Multi]

[Missing]

[DataSetYear]

[DataSetYearID]

[DataSetID]

[GridRefX]

[GridRefY]

[Year]

[DataSetMonthlyData]

[DataSetMonthlyDataID]

[DataSetYearID]

[Date]

[Value]

Specify separators/regex for data fields

Specify separators/regex for data separator fields

Phase 3

Notification of invalid data

Ability to store headers and associated regex within the database for future/dynamic parsing.

Calculate and manipulate data

Open Questions

Are all import files in the same structure?

Is a data row end value required for testing the data?