

Course Completion Report

Introduction to Data Management

Course 4.7



Macintosh HD:Users:sveinstoveland:Library:Containers:com.apple.mail:Data:Library:Mail Downloads:3FAEC3AD-6DF2-485D-9D31-E9D154D7945C:4.7 Introduction to Data Management - Course Completion Report.docx

1 CONTENT

- Course data
- Course objectives
- Training focus group
- Practical and theoretical work
- Course evaluation: comments from participants
- Recommendations for next or other courses

2 COURSE DATA:

| | |
|-------------------------|---|
| Name of course: | Course 4.7: Introduction to Data Management |
| Number of participants: | 20 trainees from: RuWatSIP, MRRD, MEW, AGS & Kabul University |
| Location for training: | RuWatSIP Conference Room, MRRD |
| Date of implementation | 1 st to 3 rd December 2014 |
| Course organiser | Mr. Andreas de Jong |
| Supporting presenter | Prof. Zarinkhail |

3 COURSE OBJECTIVES

The main objectives of this course are to (i) introduce the concept of data management and (ii) demonstrate the practical aspects of data cleansing, validation, standardization and migration using MS Excel & Access.

A prerequisite to this course is that the participants are already proficient in MS Excel & Access and/or have completed the previous two training courses in Data Management which introduced Excel & Access.

The course was designed to be mainly practical and taught using hands-on training using a variety of real data.

Why do we need Data Management?



Course Completion Report Introduction to Data Management

Course 4.7

NORPLAN 

Macintosh HD:Users:sveinstoveland:Library:Containers:com.apple.mail:Data:Library:Mail Downloads:3FAEC3AD-6DF2-485D-9D31-E9D154D7945C:4.7 Introduction to Data Management - Course Completion Report.docx

4 TRAINING FOCUS GROUP

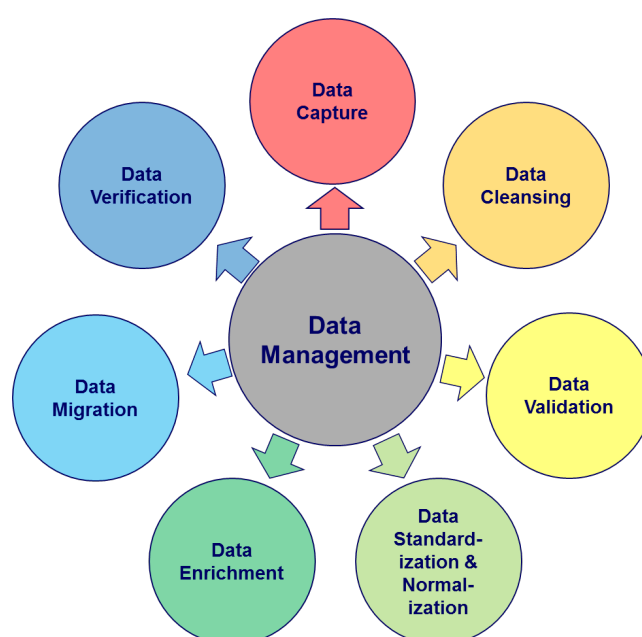
The focus group of this course are National / Provincial engineers working in the hydrogeological and GIS fields. Participants from the various ministries were as follows:

| No. | Organization | Persons | Percentage |
|-----|------------------|-----------|-------------|
| 1 | RuWatSIP/MRRD | 10 | 50% |
| 2 | MEW | 5 | 25% |
| 3 | AGS | 2 | 10% |
| 4 | Kabul University | 3 | 15 |
| | Total | 20 | 100% |

5 PRACTICAL AND THEORETICAL WORK

The course was designed to cover two days with a presentation and tutorials in MS Excel on the first day followed by tutorials in MS Access on the second day. Due to the slow progress during the first day, it was decided to extend the course to a third day.

Day 1: On the first day the participants were introduced to the key data management concepts as shown in the diagram below. Each step was illustrated by practical examples from Afghanistan and elsewhere, and the importance of good quality control was emphasized repeatedly as the key issue which participants should concentrate on to save time, money and reduce errors in their daily work.



| | | |
|---|------------|--|
| Course Completion Report Introduction to Data Management | Course 4.7 | NORPLAN  |
|---|------------|--|

Macintosh HD:Users:sveinstoveland:Library:Containers:com.apple.mail:Data:Library:Mail Downloads:3FAEC3AD-6DF2-485D-9D31-E9D154D7945C:4.7 Introduction to Data Management - Course Completion Report.docx

The first practical was a simple example of data extraction in MS Excel using very clean data from the European Central Bank. The NOK/EURO exchange rate data was processed using:

- the Text-to-Columns tool,
- the Concatenate function to stitch together the day/month/year data
- standard copy-paste and Excel functions (LEFT, MID & RIGHT).

This practical took much longer than anticipated, which just shows how much the participants are still struggling with the practical use of Excel. There were some unexpected problems with the data processing on some laptops which were fixed by setting the regional setting to UK data formats from the US settings.

Day 2: The second tutorial is a realistic example of data cleaning using some very problematic data of water levels from Africa. Text and numbers have been mixed at random in the same data columns, and there are numerous data entry errors including invisible spaces. The processing included:

- Copy – paste special to a new sheet to eliminate formatting
- Data filtering to quickly identify dirty data.
- Find and replace tool to eliminate all unwanted text and typing errors.
- Formatting data as a table in Excel.
- Importing the final table into MS access.
- Identifying data import errors in MS Access & how to find the source cells in Excel.

The third tutorial involves real data from several districts in Faryab Province. Although the data is clean, the key challenge is to create a single column header which conforms to good data management practise, so that the data can easily be exported to ESRI shapefiles and ESRI online at a later stage. This issue was discussed at length, and header naming principles and concepts were demonstrated with a class exercise.

Overall the progress was much faster, as the participants started to work at a normal speed. The participants requested to extend the course to a third day.

Day 3: The final day of the course concentrated on MS Access using our Faryab data. The following topics were covered:

- The importance of good, short field names for successful migration to other data platforms like ESRI shape files.
- The USGS 15-character coding system for asset management & how to use my Excel tool to create a unique code automatically.
- How to make a dual language database – example from Libya.
- How to look for duplicate data using duplicate queries in Access.
- The steps we took to eliminate duplicates from the data base using simple data verification techniques in Access.
- Extracting relevant data using Select Queries & checking them for duplicates.

Course Completion Report Introduction to Data Management

Course 4.7

NORPLAN 

Macintosh HD:Users:sveinstoveland:Library:Containers:com.apple.mail:Data:Library:Mail Downloads:3FAEC3AD-6DF2-485D-9D31-E9D154D7945C:4.7 Introduction to Data Management - Course Completion Report.docx

- Relationships between tables & potential pitfalls they can cause.
- Linking to data on other platforms such as dropbox.
- Exporting Access data to Excel & csv.

6 COURSE EVALUATION: COMMENTS FROM PARTICIPANTS

The course was evaluated with the three standard questions as shown below. Feedback was collected and organized in an Excel spreadsheet. The course evaluation document is available on the project website with full comments.

The evaluation questions and the participants' response in percentage were as follows:

Question 1: Did the training course meet your expectations?

Response: Completely: 65%, Partially: 35%, Not at all: 0%.

Question 2: What do you think about the overall relevance of training course? Considering the limits on your time and the topics discussed?

Response: Too Long: 15%, Just right: 65%, Too short: 20%.

Question 3: How the relevant was the training workshop to your organization or project's need?

Response: Very relevant: 70%, Somewhat relevant: 25%, Not relevant: 5%.

7 RECOMENDATIONS IF THE COURSE WAS TO BE REPEATED

1. This course was designed primarily with the MRRD hydrogeologists in mind. However very few of them have to date attended the Excel and Access courses which are urgently needed to fill key capacity weaknesses, so they could not attend this course. It is recommended to rerun the course specifically for the hydrogeologists, provided that the Excel and Access courses are re-run for them as soon as possible.
2. It is important to run this course with a group of participants who is at more or less the same technical level. The prerequisite that they should have completed the Excel & Access courses is an important one, and should be a selection criterion for future courses.
3. The practical use of Excel and Access is still a challenge for many of the participants, and specific topics should be reinforced through a series of targeted, short, practical training courses.
4. Finally, it should be noted that courses such as this can only serve as an introduction to the subject, and the only way to improve capacity in the field of IT is through hard, consistent work by each participant over many years.

Report prepared by: Andreas de Jong, International Training Expert, Norplan, December 2014.