Metropolia

Sami Kiesiläinen

Improving Data Collection and Management Process of Weather Observation

Helsinki Metropolia University of Applied Sciences
Master's Degree in Industrial Management
Master's Thesis
27 May 2011
Instructors: Marja Blomqvist, Lic. SC (Tech)
Marjatta Huhta, Dsc (Tech)

Preface

Writing this Thesis has clarified several process-related issues which existence has been somehow known in my work, but it has been difficult to explain them. One of the most important clarifications has been that, although the development projects are often marketed as technological competence, it is not the most important aspect of the task. It is very easy and simple to install few automatic weather stations, but they alone are not enough to achieve the desired benefits. The main aspect is to understand the whole process of data collection and management, and help target institutes to rebuild their own processes.

I would like to thank all the interviewees and colleagues who have helped me to improve my understanding of the process. My sincere thanks also go to my supervisors Mrs. Marja Blomqvist and Mrs. Marjatta Huhta, as well as to other teachers and supervisors in Metropolia for their lectures and support in writing this Thesis. I would also like to add my special thanks to Mrs. Zinaida Merezhinskaya for her active role in proofreading the Thesis.

Sami Kiesiläinen

Vantaa

May 8, 2011

ABSTRACT

Author Title Number of Pages Date	Sami Kiesiläinen Improving Data Collection and Management Process of Weather Observation 83 pages + 2 appendices 27 May 2011
Degree	Master's degree
Degree Programme	Degree programme in Industrial Management
Instructors	Marja Blomqvist, Lic. SC (Tech) Marjatta Huhta, Dsc (Tech)

Weather has been an important part of life since time immemorial. Presently, reliable weather services have high impact on society, although their benefits can not always be seen directly.

This Thesis studies the process of weather data collection and management in five cases representing four different countries. The research method chosen to investigate the problem is qualitative research using constructive case-study approach. The cases have been selected to represent a variety of societies, from modern countries such as Finland to developing countries like Nepal. Although all these cases share the same goal for the process, i.e. getting the weather observations from the observation station to the users and final storage, the process implementations vary considerably.

The outcome of this Thesis is a high-level process model which can be used in restructuring the current data management processes at the National Hydro-Meteorological Services. The proposed process model aims to avoid the most common pitfalls found in the cases. These pitfalls include, for example, parallel branches in data flow, separate data storages and the lack of automation. The proposed model would also allow for technical improvements, such as decreasing pass-through times, observation interval and improving data access. However, the proposal itself does not give any recommendations for technical solutions, because all parts of the process can be implemented in several different technologies.

Key words	Weather observation, process, case, data management, development aid, weather service

Table of Contents

Acronyms

AWS	Automatic Weather Station
CLDB	Climate Database
CWQC	Company Wide Quality Control
FMI	Finnish Meteorological Institute
GPRS	General Packet Radio Service
HQC	Human Quality Control
ICI	Institutional Cooperation Instrument
IMO	International Meteorological Organization
LHMS	Lithuanian Hydro-Meteorological Service
MFA	Ministry of Foreign Affairs
MSS	Message Switch Server/System
NHMS	National Hydro-Meteorological Service
RTDB	Real-Time Database
SEE	South-East Europe
TBM	Time Based Management
TCP/IP	Transfer Control Protocol / Internet Protocol
TQC	Total Quality Control
TQM	Total Quality Management
QC0	Quality Control, level 0 (at station)
QC1	Quality Control, level 1 (automatic, realtime)
QC2	Quality Control, level 2 (automatic, non-realtime)
UN/IDRS	United Nations / International Strategy for Disaster Reduction
WMO	World Meteorological Organization

List of Figures

Index of Tables

1 Introduction

The weather has been an important part of life as long as history of mankind exists and weather measurements have been done for centuries and. It is claimed that the first measurements of rainfall were reported as early as from Greece 500 B.C. These measurements were aimed to estimate the expected growth of wheat and used as basis for taxation of the farmers. These primitive measurements were made by using bowls. The first actual rain gauges were developed much later in Korea (Bellis 2011). The first early forecasts services were established in England, France, Germany and USA during 1800's to provide information about possible storms for seafarers (Hautala et al. 2007: 73). Although there is a long history of weather observations, their importance to everyday use has grown enormously during the last few decades and especially during the last ten years.

1.1 Weather Observations in Modern Society

Reliable weather services have high impact on society, although their benefits are not always seen directly. Several studies suggest that investing to weather services will benefit communities up to 8-20 times for each invested Euro. But these benefits are not only monetary. For example, according to press-release of Finnish Meteorology Institute (FMI) around 500 persons lose their lives every year due severe weather conditions in Nepal alone (FMI 2010).

A study conducted by Technical Research Centre of Finland about the effectiveness of the Finnish Meteorological Institute services, suggests that socio-economic benefits for each invested euro into the weather services provided by FMI, produce a benefit at a minimum of 5 Euros. But real benefits are likely to be even higher because this study excluded some significant sectors, such as defense and public safety. (Hautala et al. 2007)

Another report of possible benefits for SEE-countries (Albania, Bosnia and Herzegovina, Croatia, FYR Macedonia, Montenegro and Serbia) was published by the World Meteorological Organization (WMO), FMI and United Nations International Strategy for Disaster Reduction (UN/IDRS). It shows that investments into weather services will benefits these communities up to 7 Euros for each invested Euro, or even more, depending on a timescale and a country. In this report, the benefits are achieved by improving the early warning and prediction systems in the areas which are important to economics of these countries, which are very weather sensitive. These areas include, for example, agriculture, transportation, construction, tourism and general safety (Tammelin 2007).

In addition to the benefits in the medium and long range planning, weather information is crucial if people are faced with environmental hazards, such as floods or heavy thunderstorms. These kind of disasters often cause tremendous problems, especially in poorer countries. Having good weather services and warning systems is one of the steps to help nations from poverty. However, in many countries, the public image of weather services is still quite low, and their importance is not understood (Respondent A, Appendix II).

Climate change is another important topic which needs more accurate observations, although those observations do not need to be done at high intervals or received in real time. At the moment, there are several projects going on concerning climate change observations. Some of the projects are digitalizing old observations from logbooks of the weather stations and, a bit more indirectly, weather information from logbooks of ships (Oldweather 2010).

Modern technological developments of the last few decades, such as improved telecommunications, automatic weather stations (AWS) and various long range measuring systems, e.g. satellites and weather radars, have enabled a range of new possibilities for utilizing weather information. For a common person, these became evident in real time weather services tailored for individual customers, as well as the improved quality of the weather forecasts.

A typical weather service contains a number of different sources of data, which are analyzed and processed to ready-made products. These sources include traditional surface weather observations, weather radars, satellites and many others which are illustrated in Figure 1. The products of these can be forecasts, mobile weather services, aviation weather forecasts and other types of weather predictions. This Thesis focuses on the data collection and management processes of surface weather observations and hydrological measurements. Scope of the Thesis is marked with dashed line in Figure 1. Processes of other surface observations, such as solar radiation, radioactive and air quality measurements, are quite similar, so that the conclusions of the Thesis can also be applied to them. Further on in this Thesis, all these types of measurements will be referred to as weather observations. The long range measurement systems, such as weather radars and satellites are out of the scope of this Thesis.

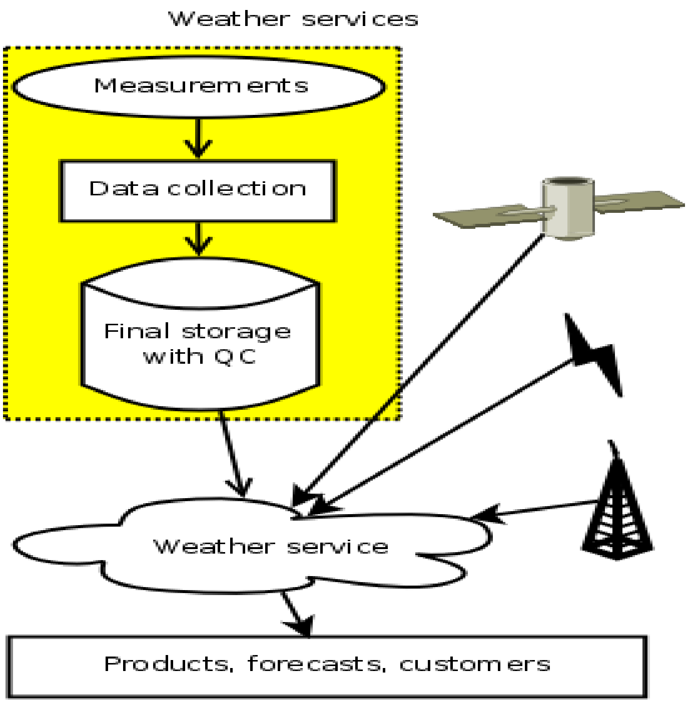


Figure 1. Different parts of the weather service. The scope of this Thesis is marked with a dash-line.

Figure 1 represents different parts of the weather service. The scope of the Thesis is marked with dash-line to figure.

1.2 Changes Caused by Automation

The field of surface weather observations has changed significantly during the last few decades. Among the major changes in this field are the introduction of the automatic weather stations; as well as new technology in sensors and data collection; and new perspectives to the data storage and management. By leaving aside the change of the actual observation methods, one of the most significant impact of automation has been a tremendous increase in the amount of data. Observation frequency has improved from 1-8 observation per day to 144 observations (with a 10 minute observation interval) or even more. Frequent measurements have also become available to users almost in real time, which is illustrated for example, in Figure 16 (page 56) and describes the situation at FMI. The significant change in the number of daily temperature observations can be seen in Figure 2.

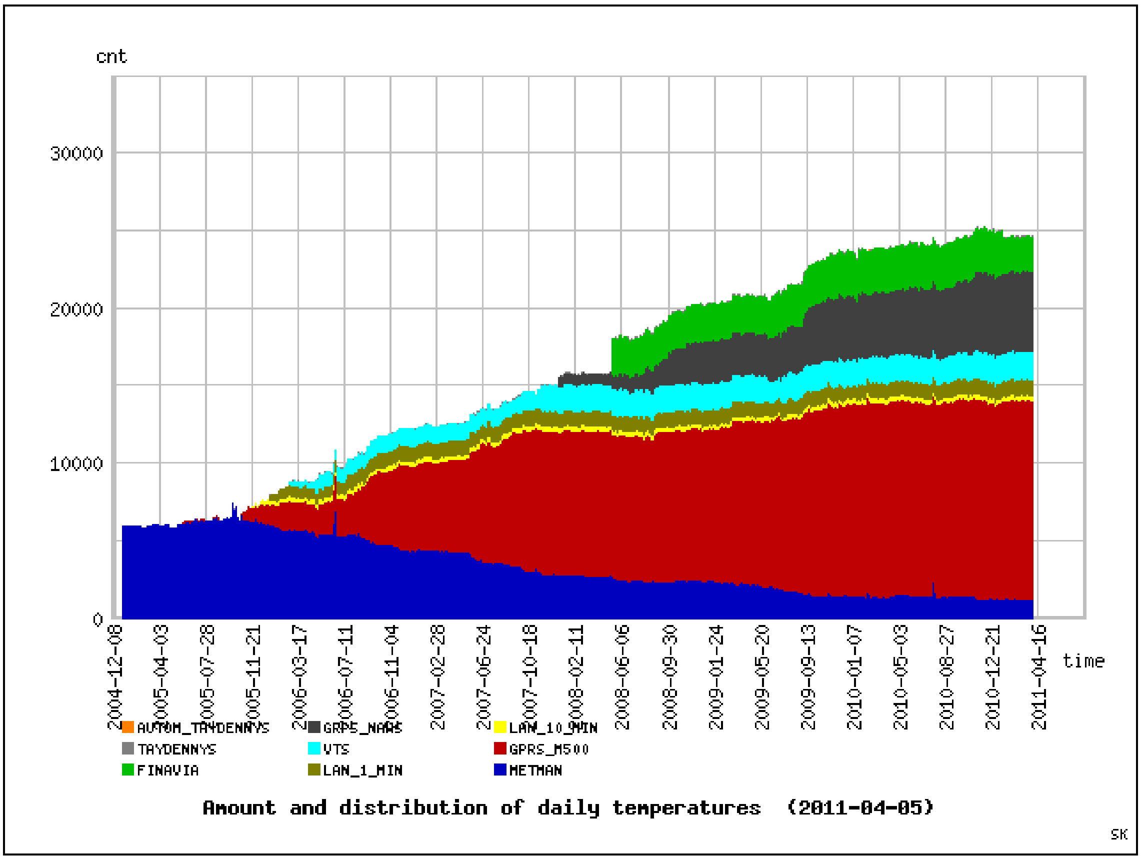


Figure 2. Growth in the number of daily temperature observations 2005-2011. Different color represent different data collection techniques. Source: internal documents of FMI.

Figure 2 shows changes in daily temperature measurements done by the Finnish Meteorology Institute between 2005 - 2010. This figure shows that the old modem connections (blue) are decreasing, but TCP/IP-based connections are increasing much faster (other colors). The amount of observations has grown from about 6000 to 25 000 observations daily.

1.3 Old Processes

However, despite the introduction of new technology at a weather station, the actual process of data management chain has not been able to keep up with the change. This can be seen as a historical burden and it is evident even in the modern National Hydro-Meteorological Services (NHMS). There are several examples for utilization of new technology suffers, or even blocked, by the old working methods, attitudes, lack of skills, organizational structures and, possibly, also by of fear of losing jobs. The present day situation is quite problematic, especially in developing countries where weather observations and forecasting have very low importance and public image, which results in low financing (Respondent A, Appendix II). These countries usually use some kind of the manual observation station network, but the data collection, data management and usage are poorly coordinated. It means that management responsibilities are divided between different departments, which makes the data collection and data processing processes very fragile and cumbersome. Very often the process has to rely on manual work at several phases, which makes it slow and error prone.

1.4 Consulting Business of FMI

Finnish Meteorological Institute (FMI) has been working since 1970's (FMI 2011) on the consulting or the cooperation projects with other countries to build up modern weather services. Finland is one of the major players in weather related development aid, second only to the USA. During the last decades FMI has been actively consulting in more than 80 countries on the different projects relating to weather services and air quality measurements. (FMI 2011) FMI also consults in this area helping other than developing countries to improve their systems and meet different regulations, e.g. by the European Union. These countries are presented in Figure 3.

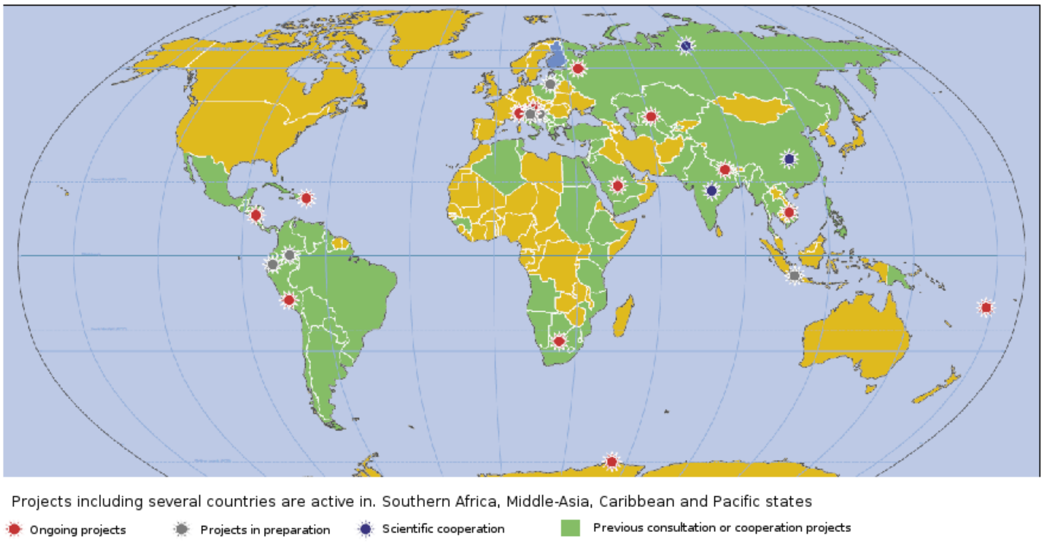


Figure 3. FMI's ongoing and planned projects, and the countries with consultation or co-operation projects (in green). Source: FMI.

Many of these projects face similar kinds of problems, and many should rethink their data management processes. The goal of FMI is to help them by providing the productized base system for data collection and management which fulfills the requirements for a modern, lean system. However, the main idea is not to provide strict instructions of how to do things, but instead to provide them with models and examples how they can develop their own systems. In addition, the data collection and management systems introduced by FMI, might work as a starting point for them to develop the processes further.

This Thesis studies the processes of the data management chain of weather observations, from the station to the database and attempts to answer the question if the current process presents a high-level model of optimal process. The research plan is shown in Figure 4 below.

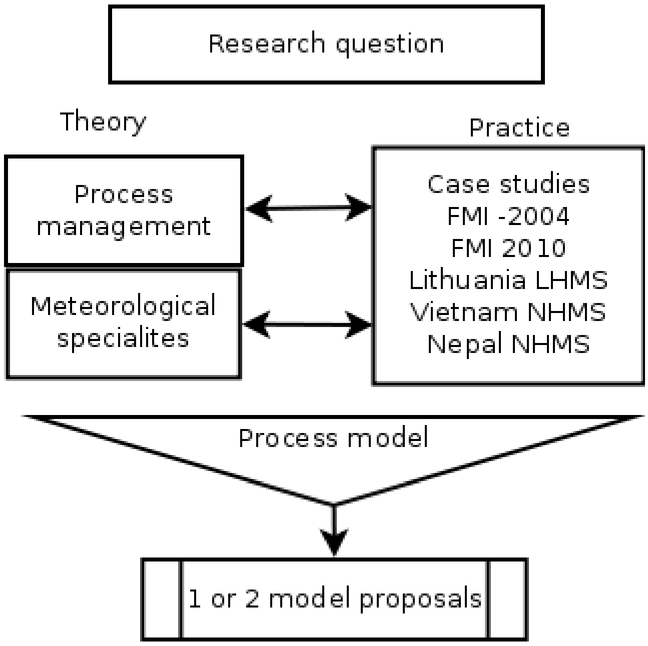


Figure 4. Research plan overview.

Figure 4 illustrates the research plan of the Thesis, which includes five cases of different models of weather data collection and management processes are studied. The goal of this research is to create one or two process models for an optimal process. These models will be useful for all organizations which have old processes in the area of weather services. However, the results can also be applied to any other similar processes where simplification is needed, including processes at FMI.

Most of the technical details are left out because different parts of the process can be implemented in several different technologies.

The Thesis investigates the process of data collection from the observation stations to the storage of the data including quality controlled database. The results can be applied also to similar measurements, for example solar radiation, radioactive measurements and similar measurements. The real real time data feed and very dense data (interval less than 1 minute) are excluded. However, it should be possible to integrate them into the model with certain small modifications.

This study also excludes the process which happens after the data is stored to the database. This phase is referred to as “production” and it includes, among others, www-pages, customer products and forecasts. That part of the process is worth of Thesis of its own.

The results will be used in the ongoing FMI's Vietnam ICI-project, as well as partially in the current Nepal ICI-project.

2 Meteorological Data Management and Collection

This Section focuses on the overview and history of weather observations. For the purpose of this Thesis, the term weather observation includes also any similar surface observation, such as hydrological measurements, radioactive measurements and all other types of surface observations.

2.1 History of Weather Measurements

Weather data has been collected and studied for decades. The oldest preserved records are from England and they date back to 13-century (Paasonen 2010). Along with development of technology, main parameters, such as air temperature and air pressure have been measured regularly from 1600s. In Finland, the longest continuous observations are from Helsinki Kaisaniemi -station which was founded at 1844 (Source: FMI station register).

Real organized weather observations and their delivery begun after war of Krim (1854 – 1856) when it was understood that it can been possible to forecast future weather and avoid damages caused by storms (Hautala et al. 2007: 73). The cooperation between countries has been managed in the United Nations World Meteorological Organization (WMO) since 1950 and before that International Meteorological Organization (IMO) since 1873. Finland has been a member of WMO from the beginning (FMI 2011b).

Old observations are used especially in climate research. Another useful source of data, are the naval logbooks of ships. Although individual logbooks exists from early times, scientifically usable collections of logbooks are dated at late 1600's (Wheeler 2005). One of the important features of the naval logbooks is that they contain observations from the open sea's. Although they do not provide all parameters with the same accuracy as the real weather stations do, these observations can be utilized. Today, there are several projects which are digitalizing old scanned logbooks for weather observations. One of the projects is Oldweather (Oldweather 2010).

Technical breakthrough in observation utilization came with new telecommunications method, telegraph, in the beginning of 1800-century. This was a start to the fast development which allowed to find out internal connections between weather phenomena in different locations. For example, this led to investigating and finding of the concepts of weather fronts by Norwegians. In Finland telegraph was taken into use during 1856. First public weather telegrams where published in newspapers at year 1881. First public forecasts in the newspapers and at the railroad stations were published 1885 (Nevalinna 2009: 21).

Another major change has been development of computers and their computing power. Also new methods of data transfer in with real time observations have brought a lot of new possibilities.

The automatic weather stations (AWS) were introduced generally around 1980. First versions were quite simple and they were overlooked by old school meteorologist for a quite long time. Currently, the AWS stations are even better than humans in areas such as wind measurements. However, there are still problematic areas for an AWS's, such as cloudiness, present weather and visibility. Although capabilities of an AWS have improved from the past, this attitude still exists.

Interesting historical details mentions some kind of the automatic weather stations in 1600 when first automatic rain gauge, tipping bucket, was created by Christopher Wren from UK (Bellin 2010). First actual AWS's were built at the beginning of 1900 and they were in used in World War II but the real growth of the AWS started as late as in the 1980s.

This Thesis focuses mainly on modern (i.e. the last few decades) observations and collection and management process during the last few decades.

2.2 Background of Data Management and Collection

During the last centuries, weather observations have been made manually, by the observation maker, reading measurement devices. Recorded values have been either stored at the station or sent a few times a year to the headquarters. With introduction of telephone, personnel at a weather station usually has telephoned them to the operator at the National Hydro-Meteorological Service (NHMS) unit which then sent them forward, improving the speed of delivery of the observations.

Even with the introduction of the telephone system, the path of the observation has been slow and difficult. Very often there has been lack of the decent tools to transfer observations, or centralized place where to store them. This has led to the situation, where the main data storage has been log-books at the station and possibly some copies in the regional center. Expected result of this kind of storage is that studies and long term research of the weather have been difficult and slow.

In cases where observations have been stored in electronic formats, there are possibilities that they are usable only at certain computer. Poor data communications and possible strict hierarchical organization structure, have not allowed cooperation between units. This kind of environment can still exists in some places.

2.3 Current Situation of Data Collection

The current situation of data collection and management varies very much around the world. In Western countries, the systems are usually fully automated and production works without human intervention. But, even in these countries, there are often historical procedures and methods in use, which can be seen as inefficient working habits, copying observations from one place to another and simulating the systems deprecated years ago. In developing countries, everything is fully manual, processes are very complex and slow and very few, if any, public products are available. In the worst case scenario, all what NHMS might produce is a booklet of yearly temperatures (Respondent A, Appendix II).

Quite often the reason for poor situation of national weather service is that the weather services are seen only as costs and even as unneeded, because "forecasts can be found in the Internet" (Respondent A, Appendix II, Popular quotation). Due to a complex and slow organization of NHMS, the potential and importance of the real time services may not obvious. Among the benefits of the working weather services are: better forecasts and early warning systems about natural hazards such as floods or thunderstorms.

2.4 Quality Control in Weather Observation Management

Quality control in any process is always an important issue. There are two different things while speaking about quality control in weather data collection and management. First one is the overall quality of the process itself, as described in process related literature (Hannus 1994, Laamanen 2001). Another quality control is quality control of the weather observation data itself. Although the latter is also loosely related to quality of the process, but it is better to handle them as different entities. This subsection focuses on the quality control of weather observations.

Theoretically, a data collection and management process can be divided into five different phases. These are: measuring, data collection, real time quality control (QC1), storage with QC2 and HQC levels, and production. In practice, there may be more or fewer phases and very often it is not possible to tell where one phase begins and another ends.

With weather observations, there are several sources for problems in data. For example, with the manual stations, most common errors are misreadings of devices and writing or typing errors (Kiesiläinen 2000: 18-19).

In the automatic stations, there is always possibility for malfunction of the devices. It is also possible to happen with the manual stations as well, but usually observation specialist can spot the problem.

With modern data communication methods, data transfers are quite trustworthy, but still some errors can occur there as well.

One issue, which is more depended on the observation methods, is that an automatic station may make too local observations. For example, when a human member of the staff measures cloudiness, they look at the whole sky. The automatic station uses a laser beam pointing upwards and calculates the cloudiness based on the reflections. This means that if there is a steady cloud just above the cloudiness meter, it can interpret the situation as fully cloudy, although there was only one small cloud. Another matter is that cloud sensors can usually reach to seven kilometers up in the sky and if the clouds are above the sensors range, AWS reports clear the sky.

A cooperation project of the Nordic countries called Nordklim has defined 4 different levels for the quality control (QC) (Vejen 2002). As a result of this cooperation, the Nordic countries, and a few other countries, are using the same terms, although technical implementations differ between these countries.

The defined levels are QC0, QC1, CQ2 and HQC (Vejen 2002). These levels have been defined during era of the telephone based data collection, so there has been small adjustments in definitions, although definitions themselves are still valid.

The first level, QC0, is done at the station. QC0 can be used only in few cases because not many of the sensors can detect malfunctions. Those which can, for example, by calculating deviation or noise between the measurements, can issue QC0-level warning about the results. Manual stations can leave a note to the logbook or report a problem directly to the observation department.

The second level, QC1, is the real time quality control, which happens as soon as observation has been received and stored to database but before it is delivered to users. This level is arguably the most important of the QC-levels, because it is the only level which must be passed before observations are delivered to users. With real time data, it is not possible for human to check all the data before it is needed. Traditionally, and by definition, QC1 has studied only data from the current station because recent observations from the nearby stations may not yet be available during the QC1.

The third level, QC2 is also automatic, but not a real time, check which can utilize also neighboring stations. Crosschecks between other stations are good when trying to find anomalies of the data. QC2 is usually done in 30 to 60 minutes after the arrival of the observation to make sure that data transfer from nearby stations has completed as well. Today, with 10-minute data, there is always recent enough data available, so many checks which are by definition QC2-checks are moving to QC1 phase.

The fourth level is HQC which means Human or Manual quality control. Definition of HQC has changed greatly from original definition (Vejen 2002), which states that HQC can be done at any place, from a station to data storage. In the current environments, HQC usually means a phase where observation quality controllers go through errors and warnings issued from the earlier levels. The amount of data is so vast, that it is not practical or possible for humans to check everything.

Each level of quality control can set flags to inform possible problems in data. Technical implementation of the flags differs in countries, but flags can be converted to be used in the other systems if needed. The flags used at FMI are 4 digit number representing different QC-levels. Each digit can have value from 0 to 9 with meanings described in Table 1.

Flag	Short description	Meaning
0	No check	No QC-checks made
1	OK	According to checks, observation is valid.
2	Small difference	Small suspicion against value but it can't be said it were wrong.
3	Big difference	Strong suspicion against value, but it can't be said it were not correct.
4	Calculated	Value has been calculated with certain equation using other values
5	Interpolated	Value has been interpolated, either by human or computer
6	Not used	-
7	Not used	-
8	Missing	Value should exists, but it is missing.
9	Deleted	Value is certainly wrong and should not be used.

Table 1. Flag values used in FMI.

The database system in FMI uses views to restrict access to deleted values. Although there is a flag named “deleted”, the actual values still exist in the database, but the users have no access to them.

3 Management of Weather Observation Process

The meteorological data collection and management process is a service process which produces many kind of the results. As defined in the scoping of the Thesis, a process is limited to observation made at measuring station to storing it to the final storage. The production phase, which produces the results for end user, is only briefly mentioned. By excluding also the research measurements, it can be said that result of the process is to have a quality controlled observation at final storage.

Many features related to the Total Quality Management (TQM) thinking, such as “no errors” idea and continuous improving, do fit well to this kind of process. Other process management models, such as the Time Based Management (TBM) often focuses more on manufacturing processes or processes where the outcome changes often.

3.1 Stability of Process

This kind of process is quite static, because measured parameters are well known and they change quite rarely. New parameters appear from time to time, but existing parameters usually stay. Major changes in parameters happen when two existing data collection and management processes are merged. This kind of integration is happening, for example, in FMI when weather observation stations are equipped with marine observation equipment.

The data collection methods do change, but slowly. There are two different ways to implement new systems, either replace the old system with new at a certain moment, or gradually by moving observation stations to the new systems. One-time change happened with the FMI's Metman system in 2004 while the old VAX-based data collection system was replaced. However, lot of equipment was preserved, because the both systems use the same modems for data collection. Bigger, but gradual, change has happened while weather stations have been changed to the TCP/IP-based data collection system. In this case, change is done one station at a time because it requires a new telecommunication device for the stations. There has been also decision that oldest AWS-stations of FMI are not be supported by the TCP/IP-system, so stations must be upgraded to new technology.

In general, reception and decoding of a message from weather station is closely related to the data collection system in use. However, with the modular systems, the part where message is acquired from the station is being separated from the parsing, (i.e. decoding) but in general, it can be said that parsing programs are quite stable as well. This phase is combined to "Data collection" box in process descriptions.

Storing observations on the data storage, intermediate or final storage, is quite stable process as well. While using intermediate storage, storing routine is very flexible and does not need modifications if new parameters appear. Depending on how final storage has been designed, new parameters may demand at least configuration changes and in worst case even changes to the program.

Automatic real time quality control, QC1, is usually integrated more or less to data storage routines. QC1 is the part which can always be developed further but changes should be made very carefully.

Manual quality control, HQC, is very stable part, after all needed programs have been created. New errors and warnings can be implemented to the earlier quality control levels, but for HQC, they work as any previously existed observation. Depending on design of the HQC program, new parameters may need modifications to program itself.

So, in summary it can be said that a data collection and management process, if it has been properly designed and implemented, is quite stable system and there should be bigger changes only if technology in some part of the process changes.

3.2 Continuous Improvement

One of the items which belong to Total Quality Management (TQM) is continuous improving (Hannus 1994: 133). It has been in use especially in Japanese models Total Quality Control (TQC) or Company-Wide Quality Control (CWQC). By studying parts listed in the previous subsection, it is possible to identify places which need improvements and what kind of possibilities for improvements there are.

Almost all of the aspects can be improved, but not all improvements are equal. For example, increasing observation interval from 3 hour to 10 minutes is much more important than increasing frequency from 10 minutes to 5 or 1 minute (Respondent F,

Appendix II). Although it is technically possible to use 5 or 1 -minute data, with the current technology, costs rise more quickly than benefits in on shorter interval. It can be said that observation frequency has a limit in which good enough results can be achieved. On some other areas, like reliability, there is not similar kind of limit. In all aspects, costs are major issue which gives boundaries for improvements.

At the observation station, reliability, both in functioning and measuring are the most important things. At a manual station, observation frequency is usually much less than in an automatic stations. If the station is automatic, it may be located to areas which are inaccessible during certain times a year or hard to reach in general. Maintenance visit by technician may be costly.

Measurement results and their accuracy are important in operational stations. Malfunctioning devices, crawling results of measurements and other problems can cause lot of work later on, while trying to identify reliable data. It may not even be possible to fix the data at all. Accurate values and long time series are needed especially in climate research. In some other areas, such as Helsinki Testbed -research campaign, it is possible to use cheaper measuring systems, because the accuracy is not so important factor as it is in other areas. However, it is very important to know where the value has originated and how reliable or accurate it is.

As a conclusion, it can be said that reliability of the operational weather station should be improved as much as possible.

Data collection part varies very much between countries. Importance of this part depends also on other parts of the process. If the data from station is used only for calculating yearly statistics, is there any need for real time data collection? However, it is safe to assume that all of the operational stations are in daily use, so the through-put time should be minimized. This includes also recollecting the missing data. For this part to be fast enough, it should be automated. Automation usually improves reliability, as it decreases possibility of typing errors or misunderstandings between humans.

In a modern data collection system, which use TCP/IP or some other real time data transfer technique, there is not much to improve in terms of the speed. However, reliability of the transfer and possible retries should be improved as much as possible.

For example, it is calculated that average reliability of TCP/IP over GPRS-network in Finland is about 95 % (Latva 2006), which means that there should be plans what to do for those missing 5 %. In practice, GPRS-network in data transfer is much more reliable, because calculations considered only individual TCP/IP packets which were not resent. Still there is room for improvements. For example, FMI has been developing communication devices which can resent the missing data automatically and possibly use multiple data communications methods in case of problems.

It can be concluded that data transfer can always be improved but the cost may rise too high.

Storing values on a database and QC1 phases must be accurate, reliable and in the modern systems they must be fast. These parts usually compose a single point of failure, which means that problem here may stop the whole process. Accuracy and correctness are crucial because if the observations are mixed it may be impossible to fix the situation afterward. This phase must be automatic and fast, because there is no possibility for human intervention if the whole through-put time is small.

Manual quality control is important part for reliability of the observations and overall quality of the observations. This part is very human oriented so education and training are the most important factors when improving results. HQC-programs have of course big role in the work, but if they are working well enough, there is not so much need to continuously improving them. If this part is wanted to be faster, it usually requires investments in to new personnel which is something that western countries are not willing to do.

3.3 Requirements by Customers

The requirements of different users for the results differ quite a lot, and adjusting the process means to balance the requirements. There are also several situations where good enough values are good enough and investing more efforts to improve them further, won't bring as much benefits as before. In these situations, it is important to identify first if there are good enough values, when good enough values have been reached and is there easy way to improve them further. In case there are not good enough values, it means that those parts can and should be improved as much as possible.

Process features	Good enough values
Through-put times	Yes
Observation frequency	Yes
Reliability	No
Automation	No

Table 2. Process parts and improvement goals.

In data collecting and management processes, major factors are reliability, through-put time, observation interval, efficiency and to some extent also flexibility. These factors are mentioned by most of the interviewees (Respondents B, C, D and E, Appendix II).

To meet the reliability requirement, it usually means automating all possible phases of the process except HQC and possible manual observations. The need for human intervention in quality control can not be eliminated in the near future and in the case of the manual observation station, there is always observation maker included. This Thesis focuses on the process itself and does not take any stand of how the observation is made in first place. However, it should be noticed that the cost of the man years in developing countries is much less than in western countries so there may not be a need to automate stations completely.

In all cases, resources set free from the data collection and management process, would be more efficiently used in developing process further, doing manual quality control and monitoring and maintaining the process. To fill the reliability requirement, the data should flow even in situations there is no person looking after the process.

The through-put time for a observation to arrive from weather station to quality controlled storage and to user is another important factor. Through-put time is very often closely related to observation interval or frequency, i.e. the time between observations. These are areas where most benefits are received via automation. In countries where processes are old, data frequency and data flow are usually low. With modern equipment and telecommunications, through-put and interval times can be easily decreased.

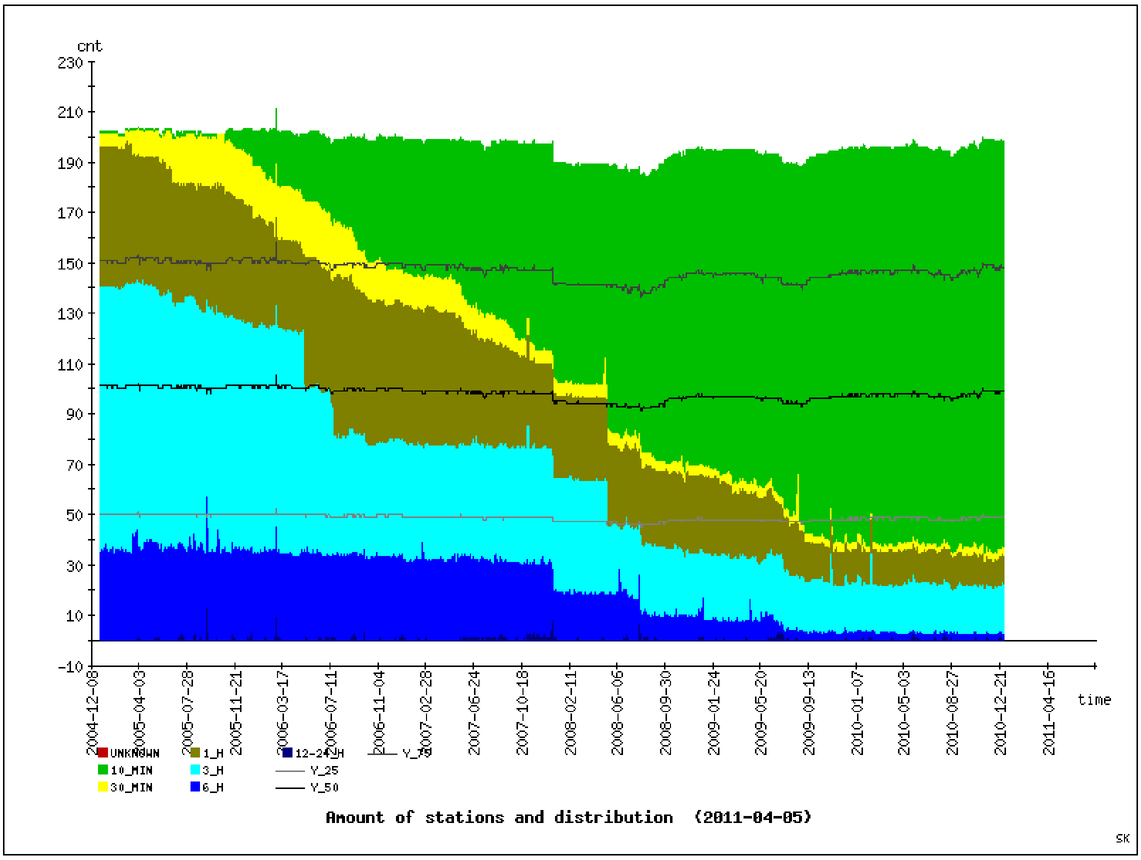


Figure 5. Change of the observation frequency from FMI weather stations 2005-2010.

However, after the good enough values are achieved, priorities of improvements should be focused on another aspects. There is a point after improving through-put times or interval do not bring relatively more benefits. According to meteorologists in FMI acceptable and good times differ depending from weather situations and usage, but it can be said that decent time for observation to arrive is 10 minutes and good interval is 10 minutes (Respondent B, Appendix II) These values are also the current goals of FMI observations. In some situations, the 5 minute data might be useful, but in most of the cases, it won't bring much more benefits. Frequency of the observations is a bit more important than their through-put time. Meteorologists are using trends for now-casting, and trend can be seen from the previous observations so improving through-put times, for example, from 10 to 5 minutes is not so big improvement than improving through-put from 1 hour to 10 minutes.

Figure 5 shows that in the beginning of 2005, only one station was sending the data every 10 minutes, and almost 70 % of the stations where sending the data every three hours. At the end of 2010, about 70 % of the stations are sending the data every 10 minutes, and 20 % have the interval of three hours or more.

Defining exact limit for good enough through-put times or frequencies of observations is difficult. Values vary depending on users, products, the current weather situation and the level of the current service. For example, in Vietnam ICI-project, FMI is implementing real time observations with 10 minute interval, although having hourly observations would be significant improvement to the current situation. Generally it can be said that the importance of the short interval between observations is growing with the ubiquity environment.

Another aspects is naturally efficiency, in cost and in other resources. Efficiency in man years is not so essential in developing countries because labor costs are cheap compared to the needed equipment. On the other hand, in Western countries labor is one of the major unwanted costs. However, maintaining reliable automatic station network is not cheap either.

Flexibility of the system is always good thing but its cost should not be too high. The estimated lifespan of a weather station is 15-20 years (Tammelin 2007). and if the system is modular enough, it can support legacy systems with small efforts. If it can not support the legacy systems, there should be enough time to replace needed parts, so the flexibility is not considered as high as other aspects.

3.4 Special Requirements of Meteorology

Total Quality Management (TQM) defines that all extra effort or resources which do not bring more value to customer are waste (Hannus 1994: 132). In traditional production or service environment, opinion of the customer is emphasized lot. In weather and climate research, long time series are important and although the official strategy of the Observational Unit of FMI acknowledges the importance of customer, there are measurements which can not be terminated even if there was only small, if any, customer interest for them. It is very possible that after few years, these measurements are highly valued. Changes in measurement programs must be considered at least in timescale of decades, not by situation of a certain year.

There is also very big diversity in customers. Research needs accurate observations from a long time period, whereas common people are more interested in the current weather and whether is it warm or cold outside.

Quality control brings new set of challenges. It is easy to find certain errors from the data, for example, having temperatures +25 ºC during wintertime in Finland. The real problems are values which are possible but unlikely, for example when temperature is approaching a new record. In this case, measurements at nearby stations do help. In cases of a short, heavy rain shower during summer, it is very much possible that there is high precipitation values in one station and no rain in neighboring stations. In some cases, the measurements from the weather radars or satellites may help, but they have their own problematic areas as well. Quality control of the meteorological data is explained in Section 2.4 (page 12).

Improving quality brings often direct costs which might be considered as unnecessary waste in short term. However, investing to quality, both in the data and in the process may save lot of work in the future. According to researchers experience, short term solutions in quality control have caused a lot of problems later which must be solved even after several years from the original incident.

3.5 Features of Optimal Data Collection and Management Process

Based on the previous subsections, features of the optimal data collection and management process of weather observations can be be summarized. Among the most important features is reliability of the process. This reliability can be achieved by streamlining the process so that there are no overlapping or parallel parts in data flow. One of the common source of problems, is having several data storages which are not synchronized. Several storages leads quickly to situation in which only part of the data has been quality controlled and possible changes or corrections are not delivered to other storages.

By implementing process without overlapping or parallel phases neither several storages, other improvements can be achieved easily. These improvements are, among others, robustness, decreased amount of errors, improved through-put times and increased amount of the data. These are not necessarily automatic benefits, but they are much more easier to implement in optimal process than in other processes. For example, by automation of certain part of the process benefits other parts as well, because their mutual order is well known. It is also possible to have different degrees of automation although most of the benefits will be achieved when the whole process is automated.

By the organizational point of view, optimal process is much more easy to manage, because data flow is straightforward, i.e. order of phases can be clearly seen. This helps a great deal to assign responsibilities of the parts of the process to proper organizational departments.

3.6 Challenges of Development

One of the major challenges in many organizations is functional thinking. Very often, especially in developing or East European countries, organization structure is very strict and everybody in the organization are interested only in their small part of the process. This results to situation where nobody has good understanding how the overall data collection and management process is working, what happens before and after their small part. In environments like this, it may be very difficult to suggest improvements, because the improvements may extend to other departments area of responsibility.

In some cases, this may be considered to be a good thing, because it allows the personnel to do only one thing which they can be trained and instructed in detail. As a result, workers can obtain high expertise in their own small area of responsibility. This point of view is common when listing benefits of the functional thinking (Laamanen 2001: 15). It was also mentioned as positive feature in Nepal (Respondent D, Appendix II). The downside for this thinking is that workers are not expected to know or understand what happens in other parts of the process, leading to inefficient methods from the point of view the of whole process. Exaggerated example could be that even if somebody is very skilled user of certain measuring device, it does not help, if he or she does not know what they are measuring.

Another example, according to our experiences in one NHMS, of the downsides of strict borders of the some kind of functional thinking, was situation where two departments were running servers in the same server room but both of them were implementing or had implemented their own monitoring systems and administrators for the servers.

Breaking this kind of the functional mindset is one of the major challenges in process renewing. It demands changes from workers as well as from the whole organization.

4 Method and Material

This Section deals with the research method and the material. The first subsection presents the research method, and the second one describes various sources of the material.

4.1 Research Method

This study is qualitative and the approach in this Thesis is constructive multi-case study. The cases used are:

- FMI before 2004
- FMI 2010 (current)
- Vietnam NHMS 2011
- Nepal NHMS 2010-2011
- Lithuanian LHMS 2006-2007

Case studies can be defined to be an empirical inquiry that investigates phenomenon in real life context at certain period when the boundaries between phenomenon and the context are not clearly evident (Yin 2003: 13-14).

The case study inquiry can manage technically distinctive situations where there are many different variables of the one result. Inquiries relies to several sources of evidence to converge data in triangulating fashion and as another result. (Yin 2003: 13-14)

Data collection and analysis of the inquiries should utilize prior development of theoretical propositions and research .(Yin 2003: 13-14)

Case study method flow can be divided in to three different phases (Yin 2003: 50):

- Planning (defining and design)
- Conducting study
- Analyzing and concluding

These phases are illustrated in Figure 6.

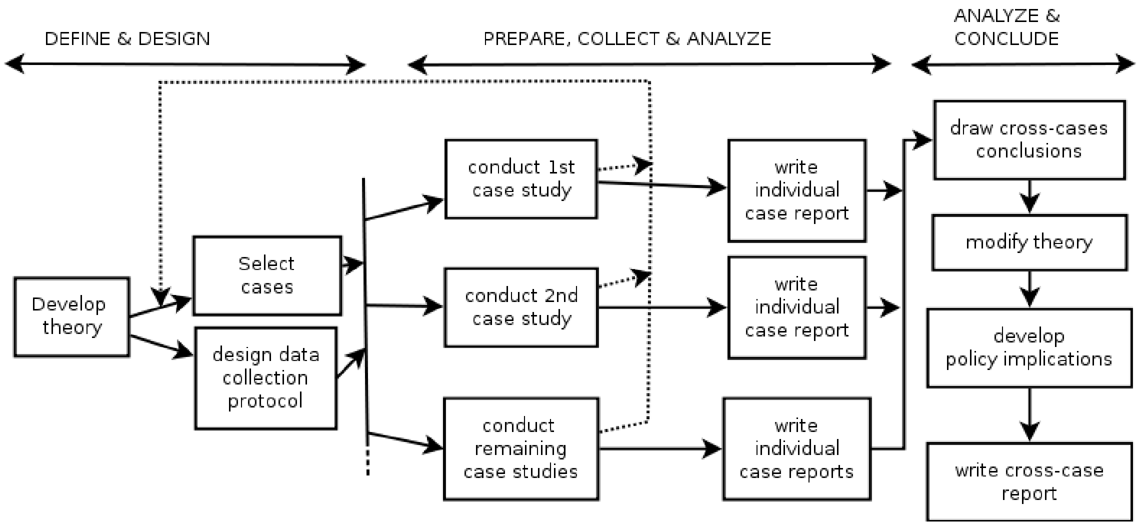


Figure 6: Case Study Method (Adapted from Yin, K. 2003: 50, Source: COSMOS Corporation)

This study follows a case study method illustrated in Figure 6 with slight differences. In "Develop theory" part, there has not been actual theory which would be studied more closely using cases. Instead of the goal for the Thesis has been studying different implementation of the weather data collection and management processes in at different countries to find a optimal process scheme.

In the planning phase, the study cases must be selected and data collection method protocol decided. These cases are selected by the availability of recent data sources and to include processes from very different kind of countries. In addition, the FMI's situation before the year 2004 is added to have more samples and to compare how the situation has changed. The Vietnam and Nepal cases are selected due ongoing ICI-projects of FMI in both countries. Researcher is a project member in the Vietnamese project and data of the Nepal case is available from the other experts in FMI. Lithuanian case was included based on the two projects carried out by FMI. Although the project was performed around years 2005-2006, the FMI experts, including the researcher, have the understanding and the documentation of the Lithuanian situation. Unfortunately, it was not possible to organize an interview on the current situation of the process in Lithuania.

The data collection protocol of these cases was either semi-structured interviews with the FMI or the local experts (FMI, Nepal, Vietnam) or documentation and experiences of the FMI's experts in the Lithuanian projects.

These case studies have been made during the research, depending on the availability of the data and the schedules of the interviewees.

The results have been compared to each other and analyzed at Sections 5 and 6. As a conclusion, a new process model has been developed, as described in Section 7. This model is suitable for places where weather data processes need to be renewed. Also, based on the results, there is mentioned another model which needs more testing before it's validity can be evaluated.

The case FMI 2004 can be considered as a pilot case study, because it was first one to be accomplished and the technical implementation is very familiar to the researcher. Main goal for the interview of FMI 2004 was to certify that all details were correct and to gain some background information.

The constructive part of the approach is to create a new model for data management process of weather observations. Constructions are defined by Kasanen et al (1993) as "entities which produce solutions to explicit problems". They also define constructive approach as a research procedure for producing constructions. The elements of the constructive research are illustrated in Figure 7.

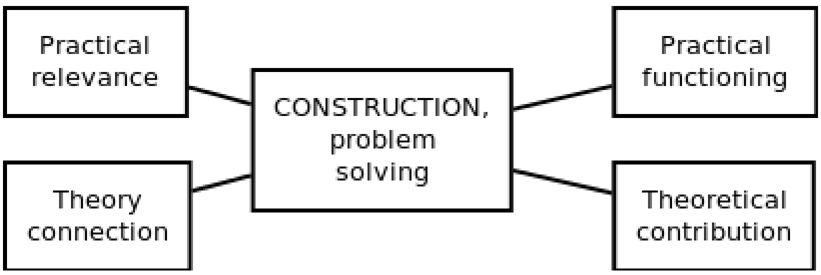


Figure 7. Elements of Constructive Research (Source: Kasanen et al. 1993).

The constructive approach is used in problem solving, when a new model or solutions are created. Examples of these kind of solutions are for example new mathematical algorithms or creating a new artificial language, such as programming language or “human spoken” , e.g. Esperanto, Braille's alphabets or Morse code. However, not all of the problem solving exercises do qualify as constructive research. Essential part of the approach is to tie the problem and its solution with accumulated theoretical knowledge (Kasanen et al. 1993).

Solution to the problem should also be demonstrated to be novelty and actually work. However, as Kasanen et al. (1993) notes, practical functioning of a construction is not at all as self evident as it may seem. Very often complex organizational processes and change resistance are present in developing and implementing a new solution. This resistance against changes has been present at all studied cases.

One of the challenges in implementing a new solution is that its benefits will be visible only after few years if related processes are renewed as well. In fact, when looking strictly from the technical point of view, the Vietnam and the Nepal projects are only required to install few automatic weather stations and a data collection system for them. This kind of set-up is very easy to do if only these requirements are considered. However, these projects bear much more bigger responsibility in a introducing new way to implement the whole data collection and management process to streamline and improve the current systems to use of the NHMS.

This kind of promotion of the new model or the working method is often so-called “invisible work”; a work which must be done but whose results can be seen only after some time. Important part in achieving good results is to change the way of thinking and adapt related processes in to the new system. If other processes continue to be the same as before, the result will be that all of the problems from the old system are preserved and one more process is added. According to our experiences in the different projects, these threats are real due to change resistance and unwillingness to transfer responsibilities and power to other departments.

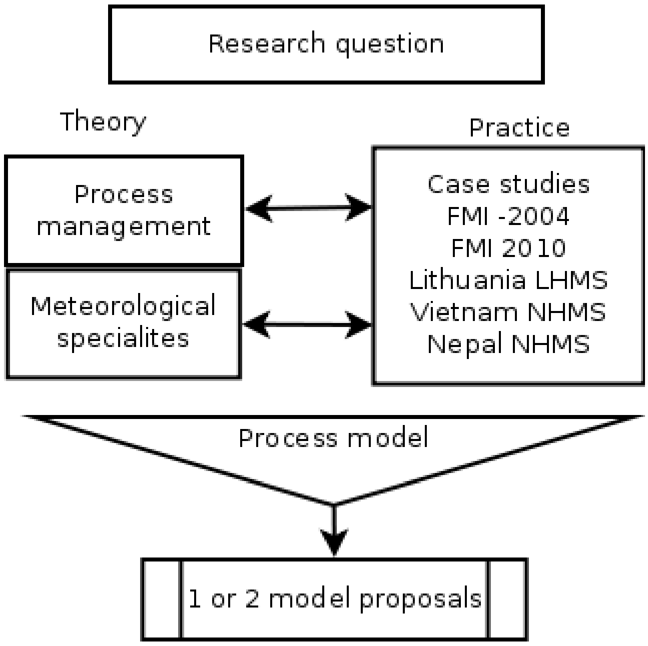


Figure 8. Research plan overview.

The research overview is introduced in Figure 8. While trying to answer research question "What kind of process model is optimal for weather data collection and management process", five case studies are concluded and they are compared to theoretical aspects of the process management, by including necessary specialties in area of Meteorology. The outcome of the study is proposal for "optimal" process model which will be implemented in the Vietnam project and so validated by the FMI and the Vietnam NHMS experts. However, the actual results and benefits of the new process model will be seen only after several years.

4.2 Research Material

To gather the research material, the structured interviews with the FMI experts have been arranged, not just for FMI descriptions, but also for identifying the quality elements of the new solution. With the Vietnamese experts several unstructured interviews were made along the project work. Information about Nepal was acquired from the FMI experts working in the project and information about Lithuania is based on the project documentation from the two projects funded by European Union and carried out by FMI. Data sources are described in detail in the next subsections. Expert interviews can be found in Appendix II.

The qualitative interviews can be divided to seven phases, according to Kvale (1996):

- Thematizing and formulating the questions
- Designing the study
- Interviewing
- Transcribing
- Analyzing
- Verification
- Reporting

In this Thesis, these phases are followed with some adaptations presented in the next subsections.

Questionnaire Used in Study

The questionnaire used in the interviews can be classified as semi structured questionnaire because, in addition to the suggested questions, the interviewees were allowed to answer freely. The questionnaire is more about process but respondents were asked to provide technical answers if possible. Technical details help to understand overall process of the weather data collection and management, although there were not strictly necessary. In cases were interviewer is not familiar with technologies used in processes, technical details may do more harm than good in disorienting the interviewer (Kananen 2008: 78). However, in this study, interviewer had quite good understanding of the topic, so all technical details can be considered as useful additions which help to understand how the process actually works.

The questions for the interviews can be found in Appendix I. These questions worked as a basis for the interview and it was not expected that all the interviewees could answer to all of them. The part of the questions in which the interviewees are asked to rate importance of certain aspects in an optimal system as well as to estimate the current situation at their system is subjective. In addition to small number of the interviews, people from different part of the process will see aspects differently so statistical methods can not be used to analyze them. However, they provide very useful information if they are analyzed in conjunction with the interviewees' position in organization.

The interviewees, topics and times can be found in Appendix II.

Selection of Interviewees

The interviewees have been selected by their role in the organization and their knowledge of some part of the processes. The researcher has quite good understanding of who are the experts in FMI cases. For the Nepal case, the logical sources for information were the FMI experts working in the project. For the Lithuanian case, no interviews were done, but the project reports and the experiences from those projects were used. For the Vietnam case, in addition to the knowledge gained from the project, the local experts were interviewed during the project related missions to Hanoi in 2010 and 2011. The interviews were carried out simultaneously but the questionnaire was not used as such. The responses were used to verify the correct understanding of functioning of the Vietnamese weather observation system.

Transcribing

Responses from the interviewees were transcribed on propositional level on paper. Using this level was natural because all the interviews were made with the experts of the certain areas and although technical details were welcome, the most important part was acquired high-level understanding of the process. In addition, with the FMI experts it is easy to check the answers later as well as making sure that notes were made correctly in first place.

4.3 Material from Five Cases

This subsection describes briefly the material from the cases and how it has been collected as well as it analyzes of validity of the cases.

Cases FMI before 2004 and FMI 2010

The data from FMI can be considered as trustworthy. It has been collected by interviewing the experts who have been working several years with the both systems. It should be noted, that the researcher of the Thesis has worked with the FMI's Metman data collection system from the beginning and has also developed new TCP/IP based listener framework for data collection. The researcher has about 10 years of experience in this field.

Case LHMS, Lithuania

The data from LHMS is based on the EU-projects LT1 from years 2004-2005 (project number: 2003/004.341/08.02.16) and LT2 (project number: 2004/016-925-03-01) from years 2006-2007. Both projects were carried out by FMI. The researcher gave a short presentation at training phase of the LT1 project and was nominated as short term expert in LT2.

Case NHMS, Vietnam

The data from Vietnam is based on the ongoing ICI project "Promoting Modernisation of Hydrometeorological Services in Việt Nam (PROMOSERV)" (MFA intervention code: 79812201), which is a mutual development aid project between FMI and Vietnam NHMS. FMI's project team of the experts has made one visit to familiarize in the current situation and to an create action plan for the project and the second visit starts the implementation new system. During the study visit, several of the personnel of NHMS were interviewed, observation locations were visited and data processes were inspected. With the second visit, a lot of details were rechecked and confirmed and first stations were installed. Before these visits, the project managers had met managers from NHMS in the project negotiations.

The researcher responsible of data collection and database parts in the project.

Figure 9: FMI experts Ha Noi Regional Centre of the Vietnam NHMS with local personnel. Photo: Jaakko Nuottokari, FMI

Figure 9 illustrates the data collecting part of the Vietnam NHMS case. NHMS personnel are describing their system to the FMI experts.

Case NHMS, Nepal

The data from Nepal is based on the interviews with the FMI experts working in the Nepal ICI-project ”Finnish Nepalese Project for Improved capability of the Governments of Nepal to respond to the increased risks related to the weather-related natural disasters caused by climate change (FNEP)” (MFA intervention code: 79811801). The project in Nepal is quite similar to the project in Vietnam, with slightly different goals. Main goal for the project is to build centralized database and to create basic calibration procedures for AWS-stations. The experts of FMI have visited at NHMS twice the during the project (situation 2011-03-11) and studied the current situation in similar way as in the Vietnam project. Nepal NHMS has also hydrological data collection process which differs very much from the weather observation process. These processes do not have anything in common, so hydrological process is excluded from this study.

4.4 Validity of Material

Validity of the material is not always easy to ensure. In this study, validity in theoretical parts is achieved by using well referenced and known sources. In the interviews, the researcher's knowledge is used in addition to the organizational positions to select correct persons from different organizations.

In practical part of the study, validity means that all of the descriptions of the systems are correct or unknown areas are at least pointed out clearly. In cases regarding the systems of FMI, validating the answers has been straightforward due the long working period in the field and quite good knowledge of the personnel and details in the area. Several details have been confirmed from personnel who have been involved to the process earlier but not necessarily today.

The projects in Lithuania were quite long, so during that time, the system was familiarized well. However, in Vietnam, the project missions have been relatively short, so all of the information acquired there, has been confirmed from several sources in adjacent missions. Information from the Nepal project relies completely to the experts of FMI.

In summary, lot of effort has been taken to ensure validity of the material. All of the sources have been checked from other sources such as the experts who have been working previously in those areas.

5 Current State Analysis of Cases

This Section studies data collection and management process in different institutes. The following conventions are used in the process graphs; In case there are several different data collection systems but they are feeding the data to the same intermediate storage (such as the FMI 2010 system including TCP/IP and Metman), there is a single data collection box. In case there are independent data collection systems which do not share the target storage, each system has its own data collection box. An example of detailed data flow of the FMI 2010 can be seen in Figure 12 (page 40). In the process scheme all these details are included into "Data collection" box.

Production phase refers to everything what happens after the final storage. It contains delivering the observations to meteorologist, forecasting models, customer products such as aviation messages, mobile weather services and www-pages.

5.1 Process in Case FMI pre 2004

Overview

The year 2004 was a turning point for the FMI data collection and management process. Installation of the third party Metman data collection software changed many processes and introduced the first version of the so called data pipe concept in the observation management. This subsection describes the situation before the Metman-system.

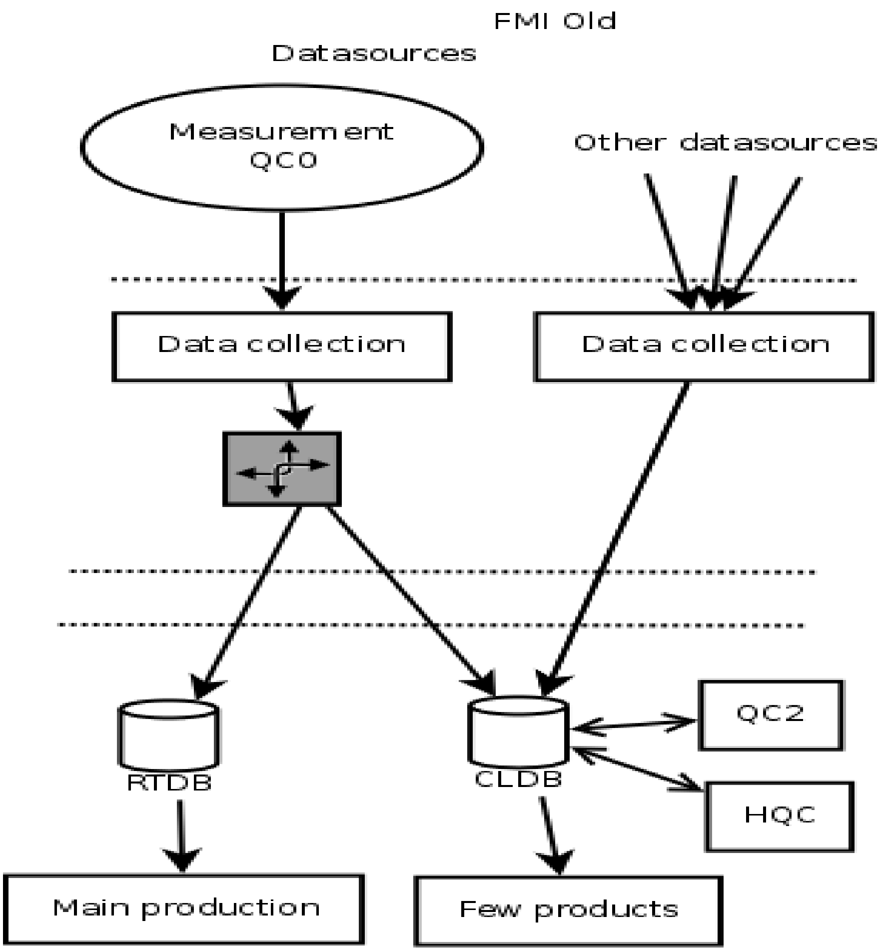


Figure 10. Process scheme of FMI before 2004.

At that time, the process of the data collection and management was quite different, compared to the current situation described in the next subsection. System included two main data storages, RealTime database (RTDB) and CLimate database (CLDB).

Advantages of FMI 2004

It is worth to notice, that already this system was automated well enough to ensure continuous dataflow without human intervention. There were several manual phases especially in production, but the basic process worked from the observation station to production. Thanks to the automated processes, pass through times of observations from station to production were decent although they were collected mostly via modems.

Challenges

The process had also some challenges, the major problem being the parallel paths starting from the data collection phase. Another problem in the main production line was that it did not have any kind of quality control, except QC0 which was done at the station. Possible diagnostic codes from AWS-stations were not used.

Challenges in CLDB-branch were different. One of them, which was not considered as a big problem, was the speed of data transfer to CLDB-database. Because CLDB was considered as a separate storage outside of the production line, data transfer was much slower. The data was loaded to CLDB few times a day and at the beginning of 2004, every 15 minutes.

Quality control of CLDB branch was in better shape than in the main line, but although there were some kind of checks done after inserting the data to CLDB, it is not obvious whether it can be called QC1 or not. By schedule, these checks would be categorized as QC2 but most of the checks were QC1 type, comparing observations only to the current station. Actual QC2 was run as an individual process. In any case, these checks were not complete and they were done only to certain part of the data, not all. The same applies to QC2 and HQC-levels.

This system suffers also from the usual problems of having several data storages. Interaction between main databases was non-existent. There was also no synchronizing between these two databases. Possible changes in CLDB were not mirrored to RTDB.

In an organizational perspective, the responsibilities of different parts of the process where scattered in to different departments, which caused some inefficiencies in developing the process as a whole.

Recommended Improvements

This system is included as one additional case study and it was already renewed at 2004, so there are no recommendations for improvements given here.

5.2 Process in Case FMI 2010

Overview

The current situation of data collection and management at FMI differs significantly from the pre-2004 situation. Major changes have happened at data collection phase, while the end part of the process has still lot of practices inherited from the last few decades. Data collection process was renewed in 2004 so that all observations goes first to intermediate storage, which is called Obsdata, because it was named by the Metman project. Currently intermediate storage is often referred as RAWDATA which is more descriptive term and Figure 11 uses also that term. This concept where a new data has only one entry point to system, is called internally as data pipe with single entrance. This means that the only point in which new data can enter to the system, is RAWDATA. The two data storages, RealTime-database (RTDB) and CLimate database (CLDB) are still there but with different order in dataflow.

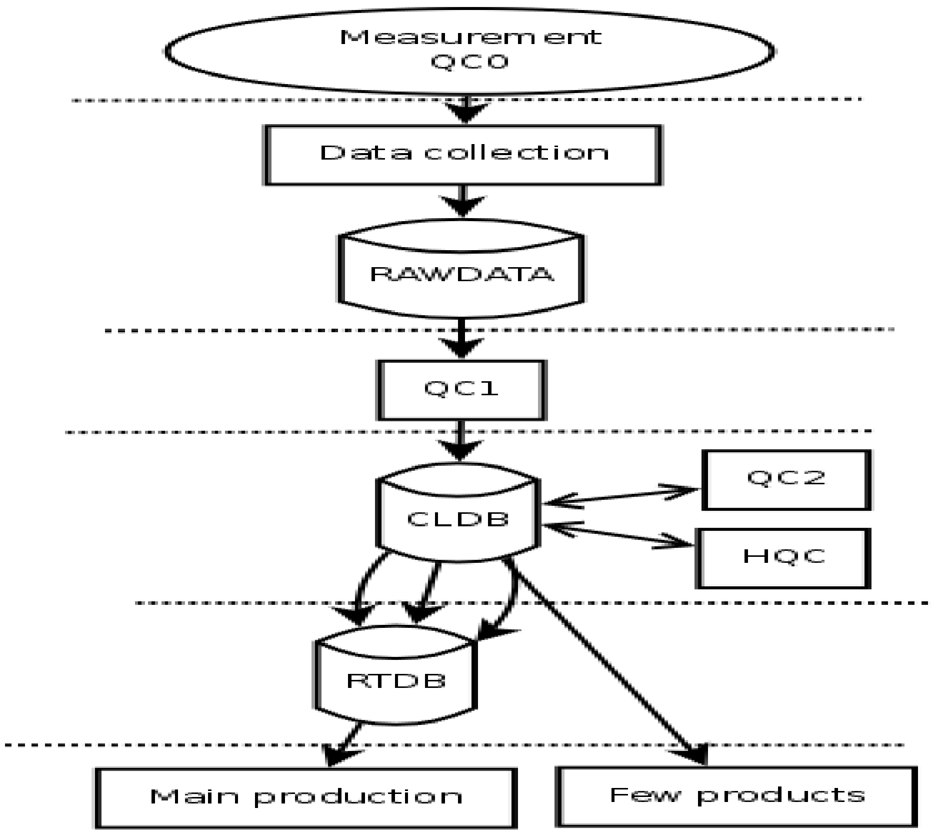


Figure 11. Process scheme of FMI 2010 data management process.

At the moment there are two main data collection systems, TCP/IP based listeners and third party Metman software, which is used nowadays for legacy measuring stations. These both systems are feeding observations to RAWDATA. These two systems are handling data collection from all surface weather stations of FMI. In addition, there are some other individual systems, for example marine observation network and radioactivity network. These will be integrated in the future to the TCP/IP-listener based collection. Air quality measurements will have, for now, their own system which will not be integrated. Also, the research projects have their own systems for collecting and managing data. The research projects can utilize techniques used in operational data collection and management process but given their very different nature and requirements, they are not going to be integrated to this operational process.

Advantages of FMI 2010

One of the main advantages of the FMI 2010 system is automation. Automation allows fast and reliable data collection and transfer to users. In addition to automation of the process, FMI has quite well working, in-house developed data collecting system based on TCP/IP communications and parser-programs. For establishing a TCP/IP connection, FMI uses mostly GPRS technique, but local area networks as well as third party networks are used as well. FMI has also studied possibilities to use radio or satellite connections. By utilizing TCP/IP-based connections, FMI has managed to improve pass-through times as well as increased amount of observations.

In addition to listener system, FMI has still some legacy systems as well. Metman data collection system is based on modems and manages mostly the old weather stations. Metman system has been connected to datapipe of FMI. Figure 12 illustrates the concept of data pipe in which stations on the left are sending data via intermediate or collection servers to RAWDATA (Obsdata in the figure).

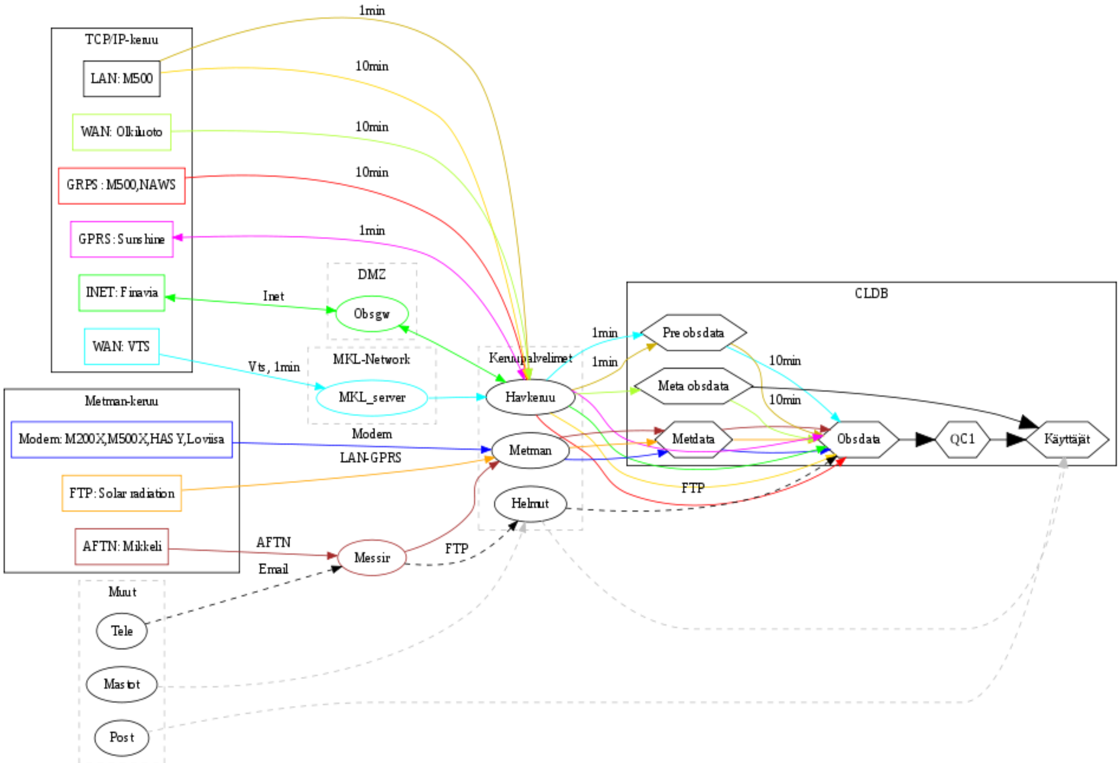


Figure 12: Detailed data collection of FMI. Colors represent different data communication methods.

Metman and the TCP/IP-collection server (havkeruu) can be found in middle of Figure 12. Figure excluded RTDB-database and production phase which are located in the process after CLDB-box.

One important difference compared to FMI 2004 system worth mentioning, is that now all of the observations must pass at least QC1 level before they are available to users. This allows to prevent erroneous observations to be used in production.

Challenges

Although FMI has a quite modern system, there is still lot of inefficient processes from the past. Anu Lintu pointed out such processes in the field of solar radiation in her Bachelor Thesis "Auringonsäteilyn mittaukset Ilmatieteen laitoksella" (Lintu 2010).

The major drawbacks of the current process scheme, is the data copying from CLDB to RTDB. Copying happens in three different ways to emulate situation about 20 years ago (1980's). Another legacy problem is that interface to RTDB expects observations to be fully manual or automatic. This is a challenge, because there are very few manual stations left and to replace them in RTDB, data from AWS is copied both to automatic and manual partitions.

Although the process has been improved so that every observation goes through at least QC1 quality control before it is delivered to users, synchronizing challenge between data storages still remains. Changes and actions done at QC2 or HQC-levels are not mirrored to RTDB.

Recommended Improvements

In the future, this copying should be eliminated and there is a plan to remove all Finnish surface weather observations from RTDB and use only CLDB as their primary source. To achieve this, several new interfaces must be developed and production routines adapted to them, but after this is accomplished, it opens several new possibilities to improve the quality of the production phase of the process.

5.3 Process in Case LHMS, Lithuania

Overview

Lithuanian Hydro-Meteorological Service (LHMS) is responsible for meteorological and hydrological measurements and forecasts. Although located in the same organization, there is not much common between meteorology and hydrology departments, so data process of the hydrological measurements is excluded from the study. However, LT1 & LT2 projects studied briefly also hydrological processes.

The data collection and management process at LHMS has lot of inheritance from Soviet Union times. They have quite good manual observation station network but no AWS, until introduced by the LT2-project during 2006. Observation management is distributed to several different departments.

In their original system, LHMS had two lines for observations, resembling more or less the situation of FMI 2004. Real time usage branch routed observations quite quickly to meteorologist on duty for usage and international delivery.

On the other line from MSS-server, observations are routed to clidata database, which has served as climatological database.

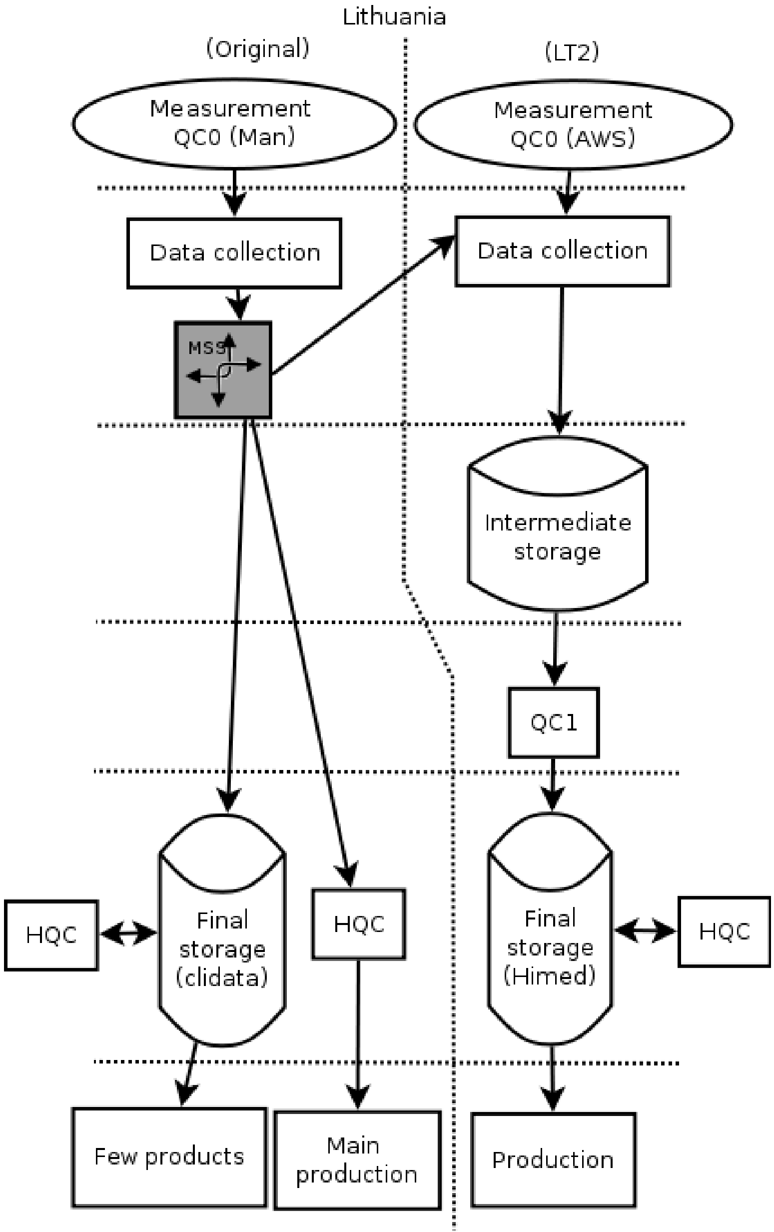


Figure 13. Lithuanian process flow with new parts introduced by the LT2 project on the right side of the figure.

As shown in Figure 13, observations from manual stations are sent to MSS server which routes them to different directions. For the ”real time” use they are sent to meteorologist on duty, who checks them before allowing them to continue. In this case, it means that observations can be sent to world and to be used at some products.

Third branch is a new dataflow process introduced by the LT2 project. It is separate from other branches although all manual observations are routed to that branch as well. Automatic observations are available only from this branch. With automatic observations, the LT2 project established a new data collection process, a new HQC as well as a new database. New database is called as Himed and it contains all weather observations from Lithuania. In the new system, QC1 and HQC are implemented but QC2 level was excluded due the limit of the resources in the project. Data collection and Himed-database are modeled after FMI 2010 system with few small differences.

Quality control is divided in to several phases and departments. It is difficult to say whether the HQC done by meteorologist on duty can be considered as HQC. It is true that person checks the synop-formatted message but they are not using other tools to verify it, resulting to the situation where only most obvious errors can be caught.

In clidata, there is decent and systematical HQC done after all observations have been arrived. However, clidata is an old system, running on old equipment and it can not meet requirements of the modern data storage.

Advantages of LHMS process

Although the Lithuanian system lacks automation from all parts of their manual observation process, their data collection and management is pretty stable. Good features are also having at least some kind of quality control in several places of the process.

Challenges

The Lithuanian system suffers a lot from usual problems of several storages as well as scattered data process. The responsibilities of different parts in the process are assigned to several departments which do cause problems and unneeded parallel work in several phases of the process. For example, manual weather observations which are sent every three hours in synop-format, are sent again each month in so called ”monthly packages” which are crafted manually. It contains more complete and accurate values and it replaces values already stored to clidata. The LHMS experts working on clidata, claim that monthly package is more accurate than observations sent every three hour. This raises naturally provocative question why to do 3 hour observations in first place at all?

One of the hidden problems of the lack of the automation is human errors while sending observation messages. Traditional synop message is very cryptic and it may be difficult to spot typing errors if it they are handmade. For example, FMI has used special computer program to create synop-messages during the last three decades. This reduces typing errors and allows to have some kind of QC0 checking. Also monthly packages are coded messages which are handmade. In discussions with the LHMS experts, they did not consider typing errors to be a real problem.

However, so far new the dataprocess is not fully utilized so old and new systems are running parallel.

Recommended Improvements

LHMS would benefit a lot of streamlining the observation collection and management process. Currently there are several overlapping phases requiring manual intervention. Appointing a cross-organizational team or teams to discuss and exchange information about different part of the process would help to spread the knowledge around LHMS. FMI has had positive experiences of having informal teams for discussion. A good example is the weather radar team which includes persons from every part of the weather radar process but so far has not had any official status or power. However, it has lot of influence power for decisions concerning radars.

Automating the most of the phases would help to eliminate or at least reduce human errors in the process.

5.4 Process in Case NHMS, Vietnam

Overview

The Vietnam NHMS has a very bureaucratic structure which affects also to the data collection and management process. NHMS is divided to 9 regional centers which are responsible of the weather stations, radars and forecasts in their region. These regions are often divided further to provinces and provincial centers have forecasting and observation responsibilities of their provinces. In addition there are top-level units for national forecasting, national IT-center and observation station network (called as Hymenet). The work flow and responsibilities between departments is quite complex. In principle, regional centers send observations to IT-center which delivers them to the national forecasting center and other users. National forecasting center makes the weather forecast which is then sent to regional and provincial centers to be improved by using local knowledge of these centers.

In Vietnam, at least in Hanoi region, everything in the data collection process is done manually. Manual observations are telephoned from the station to the regional center. The operator at the regional center receives observations as a synop formatted message. Then he or she writes them down to a text file with the help of a Notepad-program and creates a MS Word file as well. These files are sent or printed to several different locations, including national IT-center.

The original observations are stored in a logbook at the station which is sent to the regional center monthly. At the regional center, there are people who do the manual quality control of the observations but their actual methods are unclear. In any case, this quality control is not done real time.

All archives are in practice kept on logbooks and stored at the regional centers. This makes any research or usage of the data very difficult. Some of the data is stored to database maintained by IT-center and users can order data for needed stations and periods.

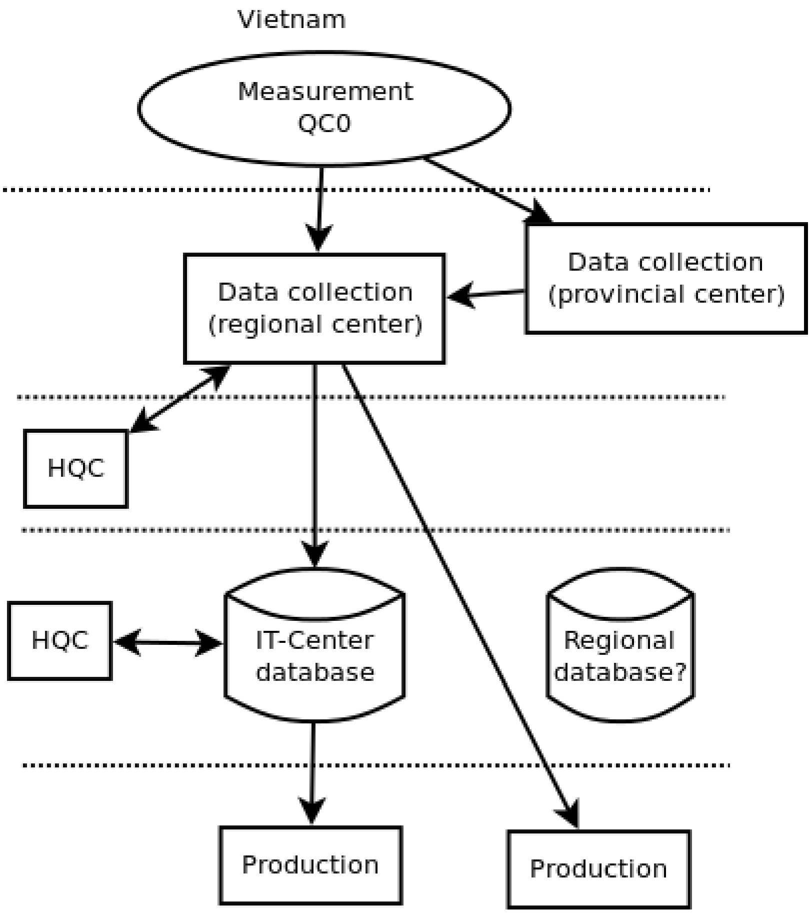


Figure 14. Process scheme of the Vietnam NHMS.

Figure 14 illustrates the best understanding of the Vietnamese data flow. However, it is unclear whether all regional centers have their own databases or not. Given the large autonomy of the regions, they have developed very different solutions.

Advantages of Vietnam NHMS process

Although the current data process includes several question marks to an outsider, it seems to work pretty well. There are also automatic rain measurement stations which can act as stand-alone warning systems at mountainous areas as well as a prototype of automatic weather station.

Personnel of NHMS have been also very interested in hearing about different systems. While a process scheme's from other cases was presented to a group of the NHMS experts they recognized that the Nepalese systems was quite near of their own in having at least two completely separated data flows.

Challenges

One of the major challenges in Vietnam, is the hierarchical structure of the NHMS and process. This makes data collection and management processes very scattered and fragile. Due to the certain features of the organizational structures, there is no single body responsible either for observation network nor for the data collection. Regional centers have lot of power to manage their systems independently, which has resulted to situations where one region may have automatic stations and another does not. Due the lack of the centralized process, only manual observations are utilized outside regions. Situations in the other regions is unknown to most of the personnel in NHMS which causes further problem in development and also the hierarchical structure requires that everything is communicated vi managers.

The scattered structure raises several questions to outsider. During the missions made by FMI, not all of the questions were answered.

Regional centers have also own meteorologists with forecasting duties, and their task is to improve the quality of forecast made at the headquarters using their local knowledge. However, the regional center has access only to observations or radars of their own region, so it is quite unclear what extra value, if any, they can bring to the forecast.

The observation stations are also under responsibility of the regional center. The regional center handles purchasing, installation and maintenance of stations. However, there is also a top level unit called Hymenet, which is responsible for approving all the equipment used at the observation station, its calibration and development of devices. Personnel of Hymenet also visit observation stations every second year, but their actual role in station maintenance is unclear.

The data process in NHMS includes several black or at least "grey" boxes and quite often individual workers do not know where the observations are coming from or even where they are going to, or how. One reason behind this is poor knowledge of IT systems. This presents a problem because even the IT-center at the headquarter was not always aware what is happening in the regional centers and what kind of connections there are.

Recommended Improvements

Streamlining the process is one of the major technical issues which should be addressed. The best way to achieve these results, might be adding more knowledge and cooperation between departments. Also, general understanding about IT should be increased as well as training the experts of certain parts, such as databases and data communications.

In the process point of view, several databases should be combined in to one main databases. Although weak IT-infrastructure sets some limits to solutions, there are technical ways to ensure that if network connection is lost, all units can continue their work independently.

5.5 Process in Case NHMS, Nepal

Overview

Data collection and management processes in Nepal are quite diverse. There are some automatic weather stations but their data is collected manually by technician from headquarters only few times during the year. This data is digitalized and stored to spreadsheets.

Manual weather stations are collected by meteorologist on duty at the airport office. Collection is done by telephoning to each stations to receive message. These messages are then stored to internal database of the forecast division.

Other observations, such as soil temperatures are collected from manual stations daily by field office. They also provide logbook of the observations every month. Field office digitalizes these observations using old computer program. All data collected by field office is sent once a year to headquarter. Headquarter crosschecks digitalized observations to logbooks but they do not have any systematic or automatic quality control procedures.

Main task of the Headquarter, is to produce statistics of the last year's weather.

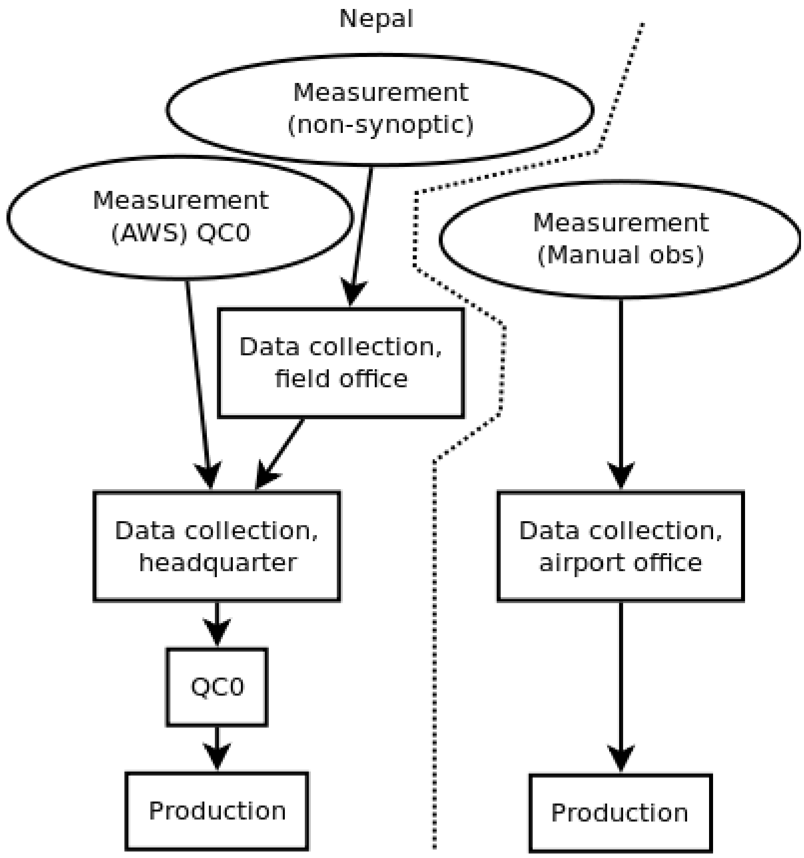


Figure 15. Process scheme of Nepal NHMS.

Figure 15 shows the situation of two independent data processes very clearly. Actually, there are three completely independent processes if the hydrological data collection was included to this Thesis.

Advantages of Nepal NHMS process

According to FMI expert working in the project, the Nepalese process is very stable and all of the personnel knows their task in the process.

Challenges

Major problems in the Nepalese systems are that forecasters at the airport office do not have access to observations of the headquarters or vice versa. They have also common problems of several data storages and with several branches of the data flow.

Data flow is very slow everywhere else except in the forecast division. Process is also lacking real quality control procedures.

Recommended Improvements

Process should be automated and automatic stations should be included to real time data collection. There should also be single storage for all of the data in order to fix the situation in which forecasters have completely separate storage from headquarters.

It can be said, and it was also noticed in the interview, that increasing automation to process would solve several of the problems.

5.6 Comparison of Cases

By comparing process descriptions of the presented cases, it can be said that there is no single model how different NHMS's have implemented their data collection and management system. It can be also assumed that functional thinking is the main guide in designing processes. It can be said that tasks and responsibilities have been assigned by hierarchical structure of the NHMS, not by process flows. Strict hierarchies have also their own share in development of the systems. If responsibilities of different parts of the processes are assigned to different departments and the experts in these departments do not know each other, it leads very often to partial optimizations which may not be good for the whole system.

Another guiding issue has clearly been the level of the IT infrastructure and knowledge. Especially in the Vietnam case, high importance of the regional and provincial centers is probably a result of the unreliable telecommunications. However, with a modern equipment and solutions these problems can be resolved.

Table 3 summarizes which part exists in different cases. Although all case NHMS's have not originally been using quality control levels defined by Norklim (Vejen et al. 2002), it is reasonable to expect that all of those levels would exists in some form or another in the process.

Final storage is defined to be "Electronic final storage" because all of the observations of the case NHMS's are stored somewhere, either on logbooks or on databases.

Logbooks are important storages as original copies of the observations, but in utilization of the observations, they have not much use.

	Measurement	Data collection	QC1	Electronic Final Storage	QC2	HQC	Production
FMI -2004	X	X	-	X	X	X	X
FMI 2010	X	X	X	X	X	X	X
Lithuania	X	X	Partial	X	-	X	X
Vietnam	X	X	-	-	-	-	X
Nepal	X	X	-	Partial	-	-	X

Table 3. Identified phases of the process.

As shown in Table 3, only few of the cases include all of the identified phases of the data management process. At some places, there are two or more parallel systems of which one includes the phases and others do not. In this situation, feature has been marked as 'Partial'.

Overall, it is obvious that all of the investigated cases do include three common phases, which are the measurement (with possible QC0), data collection and some kind of production.

6 Analysis and Results

By comparing the case studies presented in Section 5, it can be seen that processes in different countries have evolved to very different kinds of systems. In many of the cases, the current process seems to be built with one part at the time, usually maintaining old structures and procedures. This leads to very scattered and fragile systems, which should be totally restructured.

One of the reasons behind this result seems to be the strict hierarchies and functional thinking in which each department are handling only their own part of the process and does not have much knowledge of what happens before or after that.

6.1 Analysis of Identified Phases of Process

By comparing these case studies, several common phases can be identified although it may not always be possible to exactly define where one phase ends and another begins. Some cases do not include all of the phases. These phases include:

- Measurement and possible QC0
- Data collection
- QC1
- Final storage
- HQC
- Production

In addition, there may be QC2 checks.

	Measurement	Data collection	QC1	Electronic Final Storage	QC2	HQC	Production
FMI Old	X	X	-	X	X	X	X
FMI New	X	X	X	X	X	X	X
Lithuania	X	X	Partial	X	-	X	X
Vietnam	X	X	-	-	-	-	X
Nepal	X	X	-	Partial	-	-	X

Table 4. Identified phases of the process.

As shown in Table 4, only few of the cases include all of the identified phases of the data management process. At some places, there are two or more parallel systems, of which one includes the phases while the others do not. In this situation, the feature has been marked as 'Partial'.

It is obvious that all of the cases do include three common phases, which are the measurement (with possible QC0), data collection and some kind of production.

In Table 5, the most important features of different parts of the process are shown. Many of the cases studied lack automation, especially in the areas where it would be most beneficial.

	Automatic data collection	Automatic routine	Automatic production	Single final storage	QC1	HQC
FMI Old	X	X	X	-	-	Partial
FMI New	X	X	X	-	X	X
Lithuania	Partial	Partial	-	-	Partial	Partial
Vietnam	-	-	-	-	-	Partial
Nepal	-	-	-	-	-	-

Table 5. Features of different parts of the process.

Table 5 refers to the automatic fields whether there is manual work needed in the process, excluding HQC part.

6.2 Common Problems in All Processes

Based on the analysis of the cases, one of the major problem is the parallel branches of the data flow. This is usually the result of the old and complex organization. The data flow branches are usually administrated and monitored by different organizational departments which do not know, or may not even be interested, in what happens in another branch. Very often the final storage and real time usage are located in different branches, meaning that possible corrections are not synchronized at all. Separating the real time and final storages may have been acceptable solution even before, but with today's computing power and storage spaces, there is no reason for that separation anymore.

Having independent branches makes it also very difficult to improve the data management processes, because the systems from some other departments must be supported. A good example of this is the comparison of the old system of FMI (Section 5.1) and the current system of FMI (Section 5.2). Although FMI was able to streamline and improve the data collection process, there still is a problem of copying the data to several places which are not synchronized, in order to support these old solutions.

Lack of the automation of the processes can be seen clearly in process but FMI's. Automation can give a lot of benefits to the process, but it needs a totally new planning. Without automation, there is the need for humans in the data collection process. It can work, if there are only a few observation stations with long observation intervals; but as soon as the amount of observations and real time needs are growing, the human becomes outdated. With the growing amount of data, also manual human quality control of all the data will become impossible. Humans get tired while browsing a myriad amount of values from screen and may miss obvious problems. In any case, there is always possibility of the human error, such as routing or typing errors of messages.

On the other hand, humans can spot and solve some errors which are so far difficult or impossible to be handled by computers. The natural conclusion here is to invest into automatic quality control procedures, as well as to HQC part in which humans can further check the possible problems identified by automatic QC procedures.

The reasons behind these problems can be divided into organizational and technical problems. Among the organizational problems, there is lack of cooperation between the departments, unclear responsibilities; and the lack of the holistic view on the whole process. During the projects, the boundaries between the departments have been clearly seen as obstacles, well as the factors leading to the difficulty to find solutions outside of the own small area of responsibility.

Among the technical problems, there are most commonly a lack of skills, a lack of equipment and funding. If the old structures are just automated, the problems still exist and it will be harder and harder to get rid of them later. As a result, support for historical processes, old habits, protocols and methods can become a big barrier for improvements.

However, "organic growth" is the natural way how new things are implemented, and it can not be expected that all of the solutions would prove to be optional from the first try. For example, FMI has been developing its systems for decades and there still is some inefficiencies in its processes. According to one expert, the current production system in FMI is the fifth incarnation and the sixth version on the planning table (Respondent G, Appendix II). Figure 16 illustrates some of the inefficiencies which can be detected as compared to the theoretical model of the data collection and management process presented in Section 7.

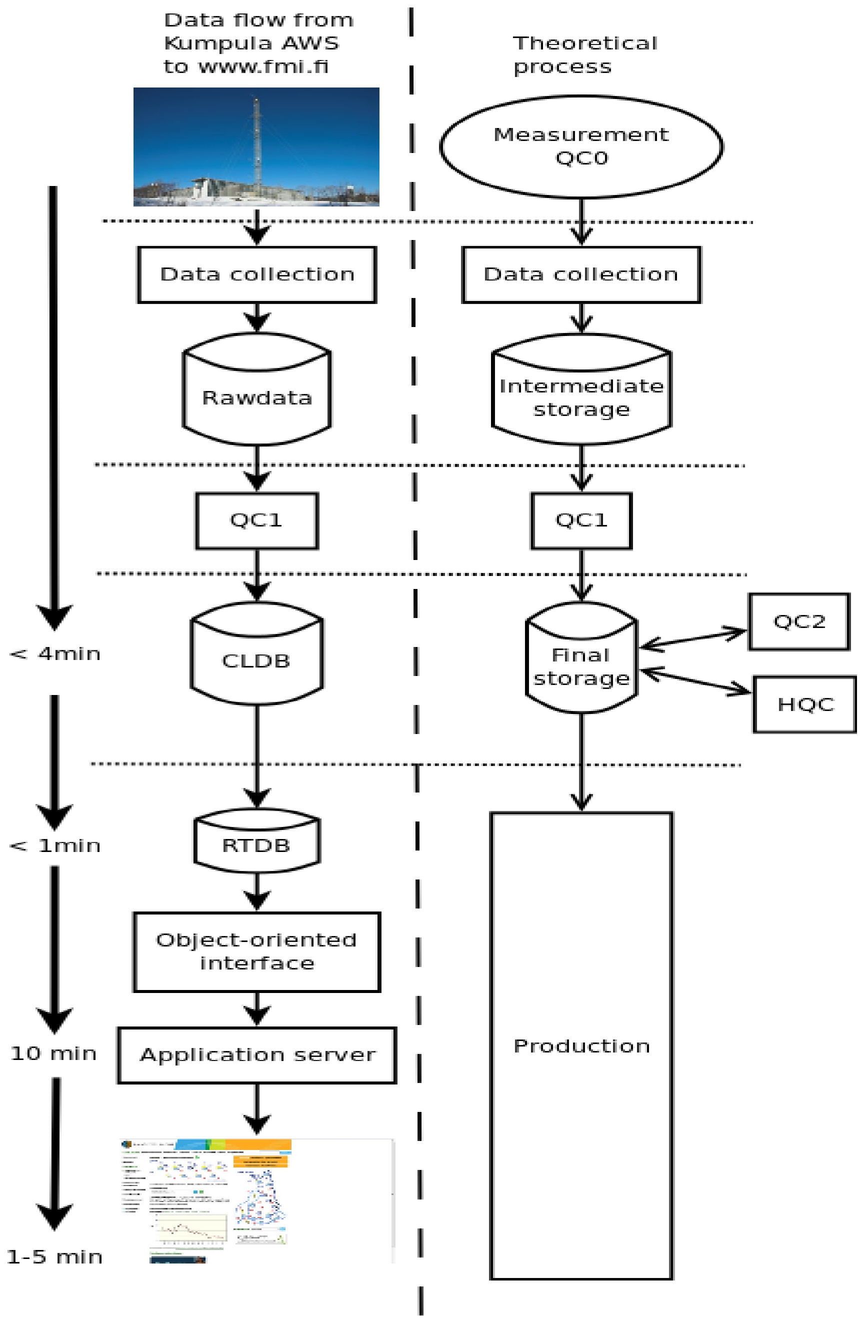


Figure 16. Data flow of Kumpula AWS to www.fmi.fi compared to theoretical process (from Section 7) with times used in process. Adapted from internal presentation of FMI.

Figure 16 presents some of the problems and delays caused by copying data in FMI. It would be relatively easy to speed up the whole process at least by 10 minutes, by eliminating intermediate steps in the production phase, but such transformation would require some major changes to the production system. From the presented cases, FMI

2010 is the only one which can be compared to the theoretical process in order to find the parts which could be improved technically. The other cases should first restructure their processes before this kind of comparison becomes feasible.

Many of the technical problems can be solved by consulting visits but organizational problems are more challenging.

As a summary, the following conclusions can be drawn out of the identified major problems.

In developing countries:

- a complex organization
- a scattered data flow
- lack of co-operation
- lack of knowledge
- lack of funding

In modern countries:

- historical burden
- old habits, protocols and methods

6.3 Summary of Cases

The most important findings from the cases are listed in Table 6.

	Advantages	Challenges	Improvements
FMI -2004	• Automatic process	• Scattered dataflow	• -
FMI 2010	• Automatic process	• Production	• New interfaces for data usage • No more copying of the values
Lithuania	• IT-infracstructure and skills exists	• Scattered dataflow • No single responsib-le body for process	• A need for cross-organi-zation teams to discuss the process
Vietnam	• Skilled people exists if they can be found.	• No single responsible body for the process • IT-infrastructure and skills • No common knowledge of the skills found from organization • De-centralized orga-nization	• A need for cross-organi-zation teams to discuss the process
Nepal	• Clear individual res-ponsibilities	• No single responsible body for process • IT outsourced	• A need for combing branches

Table 6. Analysis of the advantages, challenges and improvements from the case studies.

Table 6 illustrates most important findings on of the cases and provides some suggestions for improvements.

7 Proposed Model for Data Collection and Management Process

By analyzing the cases described in Sections 5 and 6 a range of important factors was identified to create a good weather management process. Currently none of these cases qualify as an optimal process, although "FMI 2010" is at some extent quite close. Problematic parts for FMI 2010 are data storages and production phases.

On the technical side, one of the main problems has been data flow at the parallel branches. Synchronizing data between the branches is difficult and often even impossible. However, the need for such synchronizing could be eliminated by maintaining a single storage for a certain type of data.

Another problem, which can not be defined so clearly from the process descriptions, is that, in some of the cases, new data can be added at any step, for example, at HQC phase. This again causes synchronizing problems, as well as situations in which two sets of data of the same kind, but originating from different sources, have not gone through the same quality control checks. To prevent this, all new data should be inserted to an intermediate storage (RAWDATA) by the improved data collection system.

On the organizational side, the major problem is that the responsibilities for parts of the process are scattered between different departments. This often causes situations when nobody knows how the process works and how it could possibly be improved. This is a classical example of functional thinking resulting in partial optimizations.

A possible process scheme which would resolve these problems is presented in Figure 17. The proposal itself does not take any stand as for whether the process should be located in one or several organization departments, but according to research's experience and the cases studied, the single department solution should be preferred.

7.1 Scheme for Data Collection and Management Process

The proposed scheme for data collection and management process is illustrated in Figure 17 below.

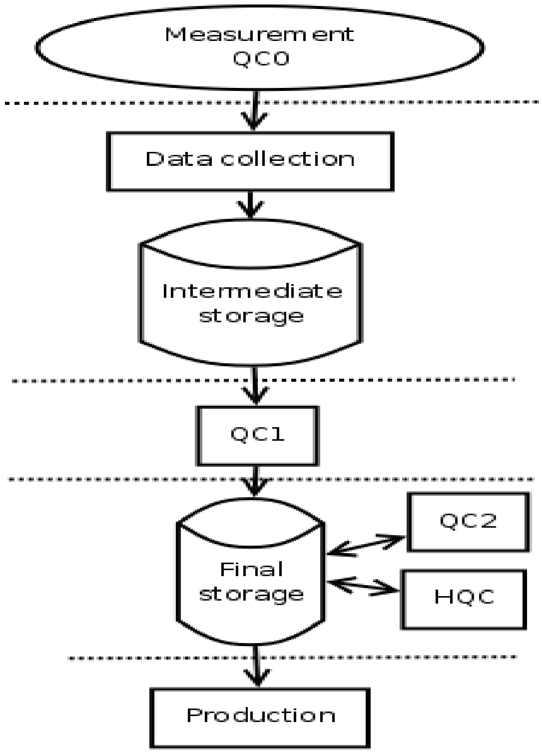


Figure 17. Proposed model for data collection and management process.

Figure 17 illustrates the proposed process model for weather data collection and management. The proposed model is quite similar to the model which FMI is approaching, although some technical details differ. In this model, data collection is separated and creates a sub-process of its own. The Data can be collected from the stations in any way needed, e.g. using a telephone, TCP/IP or web-pages but the requirement is that receiving program should know how to insert the obtained values to intermediate storage, i.e. in RAWDATA. There is no need for RAWDATA to be an actual table in database, although it is usually the case. Rawdata can be composed of several tables, and it is possible, and sometimes even recommended, that different kind of observations would have their own RAWDATAs. For example, observations from the moving stations (e.g. ships) could use their own rawdatas to include coordinates of the observation. Traditional measurement stations are mostly static, so there is no need to store coordinates with every observation. Therefore, this recommendation is dependent on the implementation of a station register and metadata information related to the stations, observations, data collection and management. QC1 process reads the data from RAWDATA , checks it, and copies it to the final database. The products are then made from the data on that database.

This kind of process ensures that all observations are available to the products and users as soon as possible, and they have passed, at least, QC1 level of the quality control. Any modifications and warnings, issued by QC2 and HQC, will be applied to all observations as soon as these phases are performed.

7.2 Alternative Scheme for Data Collection and Management Process

Another possible scheme for process might be the one in which the intermediate storage, RAWDATA, is eliminated. This kind of scheme is presented in Figure 18. Implementation of this kind of the system is not easy, because the requirements for storing the original values still exist, as well as the need for QC1, before observation is sent to users. This kind of the system will probably be tested to some extent by FMI in the upcoming years, but for this Thesis, it is left only to be briefly mentioned.

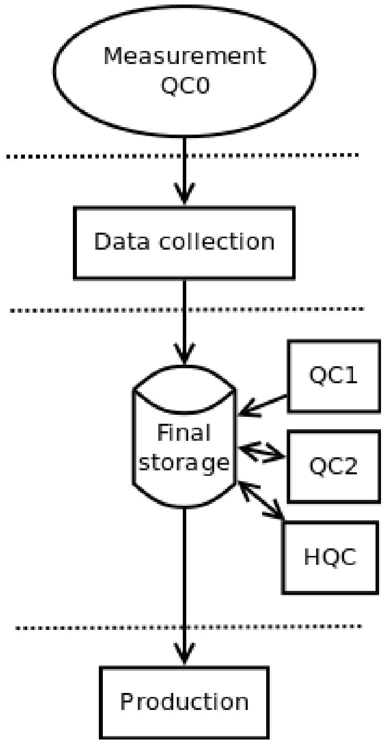


Figure 18 illustrated possible alternative for process scheme in which intermediate storage, RAWDATA, is eliminated.

7.3 Features of Proposed Model

Two of the most important features of the good process are single datastorage and single route or path for data. The proposed model eliminates the need of several datastorages which improves data quality and consistency. It also avoids parallel branches which were present in several studied cases.

In addition to these two features, introduction of "single entry point for data" concept improves consistency further. Single entry point as well as single path concepts ensures that users at production phase have access only to data which has passed at least QC1 tests.

These features allows several improvements to the process. By eliminating several parallel steps, robustness of the system increases and it often means that also amount of errors decreases.

The process model can be implemented in various ways and techniques, but streamlined process enables the full utilization of possibilities brought by automation. Advantages for automation include shorter through-put times, increased amount of data and real time quality control. By elimination of the human role in data flow, possibility for human errors, such as typing errors, decreases.

One of the additional advantages of the proposed model is modularity. Modularity allows development of individual parts of the process without disturbing others.

The features expected from the proposed model can be summarized in to the following few points:

- Data in single storage
- No parallel branches
- Automated data flow
- Single entry point for data
- All data should pass at least QC1 prior to use
- All actions and modifications to data should be stored for possible needs

It should be emphasized that human work is still needed at HQC level and there is no change to this need to be seen in the near future. Another topic worth to pointed out is that if database is mirrored to several locations, these mirrors are not considered as additional data storages, instead they are seen only as one technical solution to ensure functioning of the process.

7.4 Reliability and Validity Considerations

Reliability and validity of the study are important factors. In this kind of study, it is easy for the researcher to influence the results of the case studies with his or her own opinions. In this case, the area which most probably suffers from the researcher bias would be the conclusions where, for example, parallel processes are seen as poor. However, although it would be possible that parallel processes were useful in some situations, all of the evidence from the studied cases, as well as process theories, suggest otherwise.

Huhta (2010) defines four important topics of reliability and validity considerations. These are truth value, which is important for the internal validity; applicability, to meet requirements for external validity; consistency, both internal and external, which is a mandatory requirement for reliability; and finally neutrality, which ensures the objectivity of the study.

In ensuring these factors, the researchers long working experience, as well as discussions with colleagues have been helpful so that both good and problematic solutions in the processes were identified correctly. Building up the working process is long evolutionary journey, with a lot of trial and error situations. One of the biggest advantages of the experience has been the researcher's ability to recognize solutions which may or may not been working in some other cases. Good working experience also helps to formulate the correct questions and to interpret their validity and correctness. Correct interpretations have sometimes been quite challenging because in some cultures the experts find it difficult to admit that they do not know something or are not sure about the answer. In these cases, they may give an answer which is based on their hopes or guesses rather than facts. These kinds of answers can be identified by asking the questions in a different way again and again, and comparing answers given from different persons.

Environmental differences are not very significant topic for the reliability and validity considerations. After all, the requirements from weather observation collection and management process are quite similar everywhere. There are differences, tough, in technical details of the implementation and technical requirements. For example, the normal operation of “FMI 2010” process relies heavily on working IT-infrastructure. The countries with weaker IT-infrastructure must implement some additional measures and technical solutions to overcome these infrastructure related problems. The technical implementation may also differ a lot, but the top-level description of the process is still the same.

It should also be noted that the proposed model may not be the ultimate solution which solves all the problems. However, it is believed that modeling processes after the proposed process model would solve most, if not all, known problems in the current weather observation management processes. It is also possible that, with technology improvements, there may arise need to adjust this process model. One of the possible situations where at least re-evaluation of the model is needed, is the introduction of the real real-time and very dense data to the existing operations.

7.5 Managerial Implications

Managerial implications for the proposed model would differ depending on the organization. In FMI, development during the last years has been towards the system, which resembles quite a lot the proposed model. The main area in which the improvement work is needed, is the production part of the process, especially concerning the interfaces of the main data storage (CLDB-database). However, these interfaces and the current data delivery for production phase actually lie outside of the scope of this Thesis, but it can be said that the improvement work must also be done within this area.

For the other countries, managerial implications can be more clearly formulated. The first and most important task of the managers is to evaluate the proposed model and then compare their current processes against it. Especially the need to eliminate the overlapping or parallel processes should be emphasized. This comparison between the current and the proposed process models will probably lead, to and should, lead to reorganization of the responsibilities of the organizational departments.

This reorganization of the responsibilities may also cause a need to reorganize the organizational structure as well. However, this study does not give suggestions as for how the NHMS should be organized. Nevertheless, it can be said that several different organizational solutions have been tried in FMI. So far, the best solution has been the one, in which a department is responsible for the whole process, from making the weather observation at observation station to storing it to the quality controlled database. The quality controlled database is the first place in the process, where there is a natural border of responsibilities occurred, and it is easy for other departments to continue the process from there (marked as the production phase in diagrams). The cooperation between departments is indeed necessary, especially for the IT related tasks. Very often the department which has responsibility for data collection does not have the resources or skills to create HQC programs or administer the needed servers. It is not good to have independent IT units in different departments, because this kind of organization leads easily to weak IT infrastructure and incompatible solutions between departments without clear vision how to develop IT further. Therefore the optimal solution seems to be cooperation between IT department and other departments. In this scenario, IT department provides the infrastructure, such as technical administration of databases and data communications, and department which is responsible for data collection and management process, including algorithms used in quality control, uses those provided data communications methods and programs, and stores the collected and quality controlled observations to database.

8 Summary

This Thesis has studied different data collection and management processes of weather observations by comparing five different cases. The cases were selected so that there would be a variety of the societies represented, from modern countries such as Finland to developing countries like Nepal. Although these cases share the same goal for the process, i.e. getting the observations from the observation station to the users and final storage, the process implementations vary considerably. It is not possible to define the exact reasons for these different approaches to implementations, but they occur, at least, at the levels of IT infrastructure, knowledge and organizational issues.

According to the cases, it appears that so far there is no optimal solution found for this kind of process. This Thesis suggest a proposal for a high-level process description which would avoid the most common pitfalls found in the cases. These pitfalls include, for example, parallel branches in data flow, separate data storages and lack of automation.

The proposed model would also allow for technical improvements, such as decreasing pass-through times, observation interval and improving data access, although the proposal itself does not give any recommendations for technical solutions, because all parts of the process can be implemented in several different technologies. However, FMI provides working examples of different solutions in data collection, databases and, in the future, HQC-programs and WWW-portals in its consultation projects.

It should be emphasized that restructuring this kind of process is very complex work , which can take quite a lot of time. Although a pilot system can be implemented easily in a couple of stations, it will take a long time before all the existing systems are integrated into it. While expecting the results, it should be remembered that real benefits can become visible only after several years as improved overall quality and reliability of the process.

References

Bellis, M. (2011) About.com: Rain Gauge. [WWW document] http://inventors.about.com/od/rstartinventions/a/Rain_Gauge.htm (Accessed Apr 24, 2011)

Finnish Meteorological Institute (2010) The Finnish Meteorological Institute helps Nepal to prepare for dangerous weather events [WWW document] http://en.ilmatieteenlaitos.fi/press-release-archive?p_p_id=announcement_WAR_fmiwwwportlets&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-4&p_p_col_count=1&_announcement_WAR_fmiwwwportlets_announcementId=107936 (Accessed Apr 24,2011)

Finnish Meteorological Institute (2011a). Kansainväliset Projektit. [WWW document]http://ilmatieteenlaitos.fi/kansainvaliset-projektit (Accessed Apr 4, 2011)

Finnish Meteorological Institute (2011b). Kansainvälinen Toiminta [WWW document] http://ilmatieteenlaitos.fi/kansainvaliset-jarjestot (Accessed Apr 4, 2011)

Hannus, J. (1994). *Prosessijohtaminen: Ydinprosessien uudistaminen ja yrityksen suorituskyky*, 4th edition. Gummerus Kirjapaino Oy. Jyväskylä 1997.

Hautala, R. and Leviäkangas, P. (Eds.) (2007) *Ilmatieteen laitoksen palveluiden vaikuttavuus: Hyötyjen arviointi ja arvottaminen eri hyödyntäjätoimialoilla*. Helsinki: Edita Prima.

Huhta, M. (2010). Lecture Slides for Metropolia Industrial Management Studies: Research Skills. [Powerpoint document] (Accessed May 3, 2011).

Kananen, J. (2008) *Kvali: Kvalitatiivisen Tutkimuksen Teoria ja Käytänteet*. Jyväskylän yliopistopaino.

Kasanen, E. et al (1993). Journal of Management Accounting Research: *The Constructive Approach in Management Accounting Research*. Fall 1993.

Kiesiläinen, S. (2000) *Error Reporting System For Finnish Meteorological Institute*; Bachelor Thesis. Espoo-Vantaa Institute of Technology.

Laamanen, K. (2001). *Johda liiketoimintaa Prosessien Verkkona; ideasta käytäntöön*. Otavan kirjapaino. Keuruu 2001.

Latva, L. (2006) *Meteorologisten Havaintoasemien Tietoliikenneratkaisut*: Diplomityö. Teknillinen Korkeakoulu.

Lintu, A. (2010) *Auringonsäteilyn Mittaukset Ilmatieteen laitoksella*; Bachelor's Thesis. Metropolia Ammattikorkeakoulu.

Martinsuo, M. and Blomqvist, M. (2010) *Process Modeling for Improved Performance*. Teaching Material 2010/1. ISBN 978-952-60-3379-2 (electronic). http://lib.tkk.fi/Reports/2010/isbn9789526033792.pdf

Nevanlinna, H. (edit.) *Ilmatieteen laitos 170 Vuotta 1838-2008*. Ilmatieteen laitos. 2009. ISBN: 978-951-697-696-2.

Old Weather project (2010). Zooniverse Project. [WWW document] http://www.oldweather.org (Accessed Apr 4, 2011)

Paasonen, S. (2010) Hyvä uutinen: Sääennuste tarkentuu, Huono uutinen: Tulkinta mutkistuu (Jul 12, 2010) [WWW document]http://www.tiede.fi/artikkeli/1238/hyva_uutinen_saaennuste_tarkentuu_huono_uutinen_tu (Accessed Apr 4, 2011)

Tammelin, B. (2007) *Strengthening the Hydrometeorological Services in South Eastern Europe: South Eastern Europe Disaster Risk Migitation and Adaptation Programme*. World Meteorological Organization.

Vejen, F. (edit.), Jacobsson, C., Fredriksson, U., Moe, M., Andresen, L., Hellsten, E., Rissanen, P. Pálsdóttir, þ., Arason, þ. (2002) *Quality Control of Meteorological Observations: Automatic Methods Used in the Nordic Countries*. Norwegian Meteorological Institute.

Procurement of Services for Strengthening of Lithuanian Institutional Capacities in Ambient Air Quality Modelling and Forecasting. *Project No. 2004/016-925-03-01 (LT2)*

Wheeler, D. (2005) British Naval Logbooks from the Late Seventeenth Century: New climatic information from old sources. Sunderland: University of Sunderland.

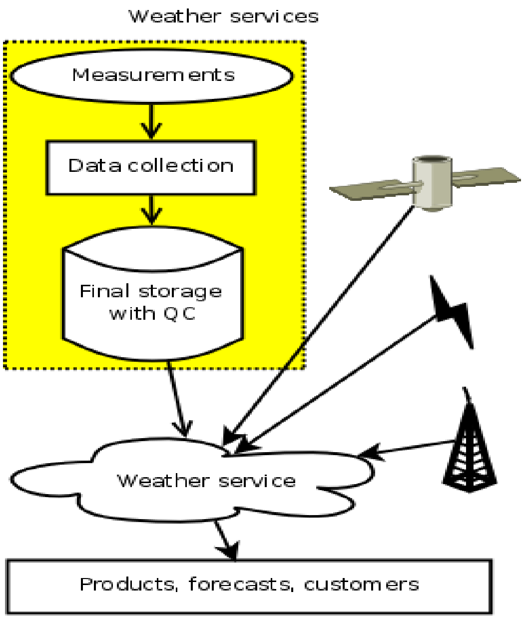
Yin, R. (2003) *Case Study Research – Design and Methods*. 3rd ed. Thousand Oaks, CA: Sage Publications.

APPENDIX I – Interview questions

This 4-page long questionnaire was used as basis for interviews.

QUESTIONNAIRE (Interview)

Your institute or service performs many kinds of measurements. This questionnaire is part of the Master's Thesis by Sami Kiesiläinen, FMI. The Thesis studies the weather data collection and data management processes in different organizations. The processes studied begins from observation station and ends to final data storage. Other processes of working weather service and what happens after final storage are out of the scope of this Thesis. Please describe how your organization manages these issues. Questions are mostly in general level, but feel free to give technical answers as well.



Current process in general

- In your current data collection and management process, what are the main
  - strengths
  - challenges

Please rate following features of optimal data collection and management process. Please also prioritize at maximum 3 of the features which should be improved in your system (if there is anything to improve).

Feature	Optimal system (scale 1-5)	Your system (scale 1-5)	Priorities of possible improvements (max 3)
Reliability (as amount of problems)			
Quality (of data)			
Quality control (checks)			
Throughput time (from station to final, quality controlled storage)			
Throughput time (from station to user)			
Observation frequency			
Cost efficiency			
Flexibility (new technology)			
Centralized data management			
Interface			
Other, what?			

Data storage

- Do you have several data storages for observation data?
- Please list your main data storages

  Storage Type of storage (database, paper copy)

Name of the storage	Type of storage (Database, paper-copies..)	Accessibility (www-pages, sql, request..)	Main usergroups of the storage (forecasters, research..)

- Do these storages contain the same data?
  - If so, how these data storages are synchronized?

Data access and usage

- If you need to access the data in data storage:
  - what actions you need to take?
  - How long does it take to receive needed data?
  - Can you get data in electronic format?
  - If you want to compare observations, e.g. say temperatures from last year, at every station in your country, can you get all needed data from one place?
  - How does the previous procedure for data access differ from accessing observations from single station?
- How is the quality control of measurements organized?

Data collection, management and quality control

- How many different units/departments are responsible for the process? Who is responsible for the whole process?
- Is there any cooperation between the units? How is this cooperation between the units organized?
- Which parts of the process are manual, not automated? How many manual (not automated) parts there are in the process?
- How long does it takes for the observation to go through the whole process from station to forecasters in the normal operation mode?
- How long does it take before this observation is stored in the final storage? (including the quality control)?

Future

- What kind of development do you see in future?
- Which things should be improved first?

APPENDIX II – List of interviewees

Group	ID	Function	Interview date	Interview type	Questionnaire response	Subject area covered
B	A	International projects	Mar 3, 2011	Private discussions	No	Importance and situation of weather services in developing countries
B	B	Aviation weather	Mar 15, 2011	Private discussions	No	Observations, reliability, real time delivery, frequency in forecasting and aviation weather
S	C	Production manager		Interview	Yes	Data collection and management process at FMI and in general
S	D	Expert in Nepal-project	Mar 10, 2011	Interview	Yes	Case Nepal General information
S	E		Mar 10, 2011	Interview	Yes	FMI -2004 FMI 2010 General background information
B	F	Commercial unit of FMI	Mar 28, 2011	Email	No	General background information
B	G		Several occasions	Private discussions	No	General background information
S	H	Project manager	2010/2011	Interview	No	Vietnam
S	I	Head of Section	2010/2011	Interview	No	Vietnam

Table: Expert interviews (B: Background information, S: Specific information)