Centre for Geo-Information

Thesis – GIRS-2008-04

Improvement of hiking network for GPS users

Laure Roupioz
840614-712-080

17/03/2008

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Improvement of hiking network for GPS users

Laure Roupioz

Registration number 84 06 14 712 080

Supervisor:

Dr. Ir. R. van Lammeren

A thesis submitted in partial fulfillment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

February, 2008
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2008-04
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Preface

It is already the end, the achievement of my studies and especially of one and half year in Wageningen University. This latter was the most fulfilling experience I had, regarding the professional as much as the personal side. Then, I would like to thank the people who participate in it. First of all, the professors, teachers, lectures and assistants, who made this master a unique learning experience, by their availability, accessibility and enlightening contact. Secondly, I thank all the GIS students, with who I worked or not, for the excellent studying atmosphere that we succeed to create.

Concerning this thesis work, I would like to thank especially my thesis supervisor Dr. Ir. Ron van Lammeren for his support, encourage and assist in my work. I would like to thank John Stuiver, who advised me to find my topic and helped me to deal with network issues. Thank you to Aldo Bergsma, Philip Wenting and Frans Rip for their time. Thank you also to Dirk-Jan, Sam, Sidney, Elena and Maria-Fernanda who were willing to help me on the field.

Finally thank you to everyone who helps me in a way or another, during this master and this thesis work. For all, Merci beaucoup!

Table of content

I. Introduction 1
1.1 Background 2
1.1.1 Global Positioning System 3
1.1.2 Pedestrian navigation 3
1.2 Hiking network 4
1.2.1 The user 4
1.2.2 Hiking network's path 4
1.2.3 GPS device 5
1.2.4 The data 5
1.3 Research statement 5
1.3.1 Open mapping approach 5
1.3.2 Research objectives 6

II. Methodology 8
2.1 Methodology approach 8
2.1.1 Hiking data 9
2.1.2 Network and segmentation 10
2.2 Study area and field tests 11
2.2.1 Study area 11
2.2.2. Field test 11
2.3 Data analysis 14
2.3.1 Existing data 14
2.3.2 Data accuracy assessment 15

III. Available Hiking route data 18
3.1 Available data overview 18
3.1.1 Web search 18
3.1.2 Data typology 19
3.2 Data assessment 21
3.2.1 Data availability 21
3.2.2 Data accessibility 23
3.2.3 Assessment conclusion 24
3.3 Data compatibility 24
3.3.1 Data selection 24
3.3.2 Compatibility assessment result 25
3.4 Conclusion 28

IV. Data collection methodology 29
4.1 Field test: sampling collection 29
4.2 Data collection results 31
4.2.1 GPS users protocol: waypoints or tracks? 31
4.2.2 Waypoints collection setting 34
4.2.3 Tracking collection settings 35
4.2.4 Collection method recommendations 37
4.3 Conclusion 38

V. Network development 40
5.1 Data overview 40
5.1.1 Data summary 40
5.1.2 Top10 vector data 40
5.2 Data processing 41
5.2.1 Conceptual processing model 42
5.2.2 Logical model 43
5.2.3 Implementation model: Application in the Veluwe area 44
5.3 Network segmentation 45
5.3.1 Descriptive attributes 45
5.3.2 Network building 47
5.4 Network validation 49
5.4.1 Path computation 49
5.4.2 Field validation 49
5.5 Conclusion 50

VI. Conclusion, discussion and recommendation 51
6.1 Conclusion 51
6.2 Discussion 52
6.2.1 Main discussion points 52
6.2.2 Comparison with French example 54
6.3 Recommendation 55

References 57

Appendix 1: Study area 60
Appendix 2: New settings for linear regression 61
Appendix 3: Implementation model in the Veluwe area 62
Appendix 4: Print screen of Georando application 63
Appendix 5: Difficulty level integration 64

Table of figures

• **Figure 1.1**: Hiking navigation components..2
• **Figure 1.2:** A Location Based Service delivering a map of the environment and the position of the hiker (Steiniger and al., 2006)……………………………………………………………………3
• **Figure 2.1:** Followed methodology……………………………………………………………...8
• **Figure 2.2:** 1st and 2nd steps of the methodology……………………………………………..9
• **Figure 2.3:** 2nd and 3rd steps of the methodology……………………………………………10
• **Figure 2.4:** The study area, Nationaal Park de Hoge Veluwe, between Ede and Apeldoorn, in the province of Gelderland……………………………………………………………………..11
• **Figure 2.5:** Fieldwork walking path…………………………………………..………………12
• **Figure 2.6:** Data collection scheme……………………………………………………………12
• **Figure 2.7:** Mean deviation computation……………………………………………………...15
• **Figure 2.8:** Smoothness scale…………………………………………………………..……….16
• **Figure 3.1:** Number of results according to search terms..18
• **Figures 3.2a and 3.2b:** Search hits by data type with and without “Netherlands” term.......19
• **Figure 3.3:** Example of digital map. 19
• **Figure 3.4:** Network of the selected existing data...25
• **Figure 3.5:** Misguidance in Otterlo..26
• **Figure 4.1:** Display of the track collected with the turn method compared to the reference track...31
• **Figure 4.2:** Top10 deviation according to distance setting..37
• **Figure 4.3:** Top10 deviation according to time setting..37
• **Figure 4.4:** Difference between collection setting based on time and distance compared to the reference route digitalised from top10..38
• **Figure 5.1 :** Footpath features from top10..41
• **Figure 5.2:** Conceptual model flowchart...42
• **Figure 5.3:** Logical model flowchart..44
• **Figure 5.4:** Example of attributes describing the GPS web offered data.............................45
• **Figure 5.5:** Network attribute table..47
• **Figure 5.6:** Segmentation level..47
• **Figure 5.7:** Network displayed according to the municipality segmentation........................48
• **Figure 5.8:** Scheme of the hiking network building rules...48
• **Figure 5.9:** Automatic computed route...49
• **Figure 6.1:** Walking path information displayed on the top of the IGN scan25...................54
• **Figure 6.2:** Examples of hiking route visualization in Georando...54

Table of tables

- **Table 2.1:** Fieldwork data collection plan.....13
- **Table 3.1:** Data availability criteria summary.....22
- **Table 3.2:** Data availability criteria summary.....23
- **Table 3.3:** Existing data point density.....26
- **Table 3.4:** Offered GPS data mean deviation.....27
- **Table 4.1:** New collected data sampling.....30
- **Table 4.2:** Garmin waypoints evaluation.....32
- **Table 4.3:** Garmin tracks evaluation (with default settings).....33
- **Table 4.4:** Bluetooth GPS+PDA tracks evaluation (with default settings).....33
- **Table 4.5:** Garmin data cross-analysis.....34
- **Table 4.6:** Pearson correlation analysis table.....35
- **Table 4.7:** Pearson correlation analysis table, with settings factor detailed.....36
- **Table 5.1:** Integration effect on the mean deviation.....45

Abstract

Nowadays, there is an increasing use of GPS devices for pedestrian navigation. A broad diversity of data is offered via Internet but they are widely spread and poorly described. From those observations, there is a need to develop the hiking network. This thesis aims to improve hiking network for GPS users through the establishment of a suitable methodology based on an open mapping approach. This choice generates the use of existing data sets and the collection of new ones with GPS devices in order to keep the method cheap and technically accessible.

The research explains that a lot of data are available, but not all are suitable for use. The first part of the thesis explores the data offered via the web and checks their usability. The performed web search offers a data description and classification as GPS data, digital maps or geo-data. A suitability assessment based on accessibility, availability and compatibility criteria leads to the conclusion that GPS data presents a broad heterogeneity and limited metadata, but appears to be suitable for a network application.

However not all real world walking paths are available in the format of data. Then, finding the most suitable way to collect data with GPS devices is necessary to enrich existing data sets into a network. The second part of the thesis presents techniques (waypoint or tracking), and settings to collect data sets and several tests to ensure the data collection method efficiency. Fieldwork analysis leads to the conclusion that using the tracking collection mode with a setting of one point every 150 meters is the most convenient. However, the results point out that it this number is more an indication than a strict requirement because it also depends on other factors, such as the nature of the device and the surrounding landscape

Finally, the last part of the thesis presents the network construction out of the suitable existing data and the collected data. In order to consider a broader scope for the research, footpath features of the Dutch Topographic dataset are integrated in this process. Then, relevant attributes are selected and the network segmentation is adapted. This latter is based on the location (municipalities and provinces boundaries) and land use category (build-up, agricultural and natural area). The result is a network created for a study area in the Veluwe and which has been validated in the field.

The offered methodology gives the guidelines to improve a hiking network for GPS users from the open mapping point of view. This study brings a sound starting point of further development to keep on improving the hiking network for GPS user.

Key words: GPS, waypoint, track, network, segmentation, difficulty level, hiking data, pedestrian navigation.

I. Introduction

After a brief introduction, this chapter is going through a presentation of the background which leads to consider the current issue. Then, some aspects about the hiking network and its characteristics are discussed. From this, the research statements are established. It describes the chosen approach and defines the research objectives, in relation to the report outline.

Since humans can walk, they try to find relevant ways to orientate and navigate themselves. To do this, they started by referring to natural elements like stars. Then they used drawings which developed and evolved into maps. Those latter kept improving to become digital and interactive. Some tools were also used to help the navigation, for example the compass and more recently Global Positioning System (GPS) devices. Originally walking was the main mean of transportation. Nowadays it tends to be replaced by others, like car or bike, and to become more and more popular as a leisure activity. However, the need of being guided remains.

At the same time, outdoor GPS equipment has modernised, gained flexibility and lost weight. It explains the increasing popularity of the pedestrian navigation with GPS devices. On the other hand, information about walking paths is still very diverse in terms of source, structure, availability and language. It is difficult for walkers to get all the hiking path data they need, even if much data is already offered (*Walkonweb, 2007*).In fact, as if the road and partially the bike networks are well known and implemented in GPS navigation systems, there is little cohesion in data concerning the walkways, which constitute the main network used by tourist out of the urban area. Then an actual need of standardized data (*FGDC, 2008*) and extending the existing network is expressed (*Arlington, 2006*). This results in the following research question:

"How to create a suitable digital network and then extend the existing ones for tourists and hiking applications?"

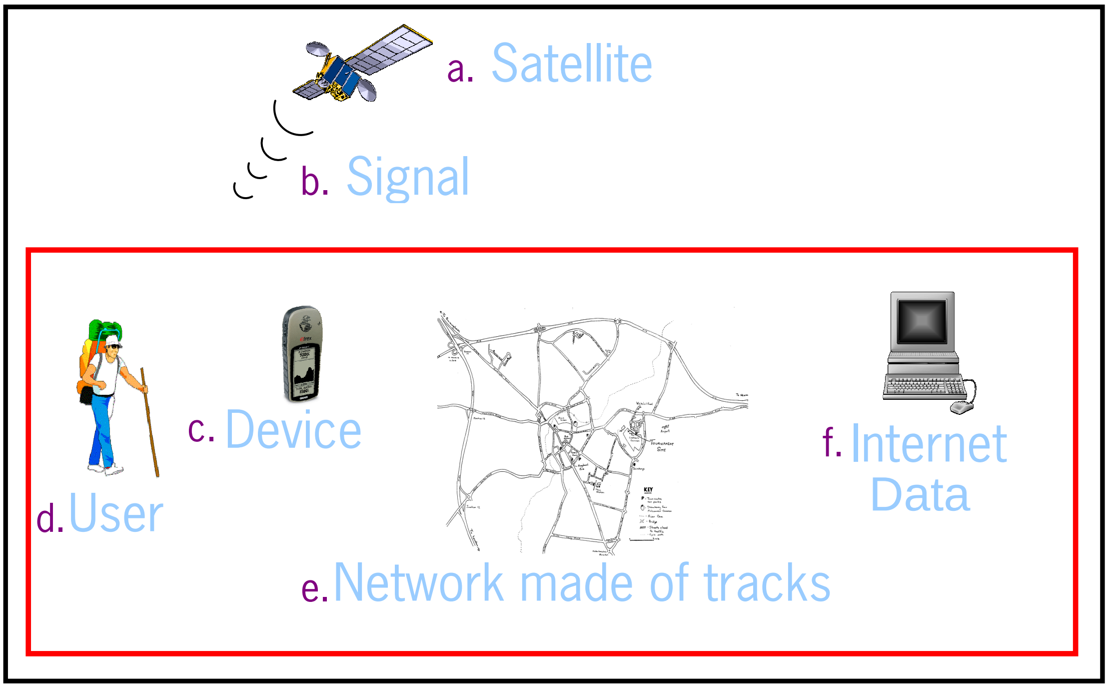


Figure 1.1: Hiking navigation components.

When talking about hiking navigation, it refers to different components, as presented in figure 1.1. First component: the GPS technology, namely a signal [b] transmitted between satellites [a] to a receiver [c], used to navigate. Second component: a network [e] made of tracks that are the basis information used for the navigation. In this case, a part of the information can already be found via the Internet [f], which is then a component also. Those components are related by the walker [d] who will use GPS technology to navigate through the network and use the data.

This project focuses on GPS applied to pedestrian navigation and especially to hiker purposes. It deals with user [d], device [c], network [e] and data [f] interactions, represented by the red square in figure 1.1. The questions concerning the GPS principle and signal transmission are beyond the scope of the study. Concerning the data, this research deals with digital data only, excluded all the data that can be found printed or published, because the network is meant to be used by GPS users. Other types of data, like booklets, are not of interest here.

1.1 Background

Before proceeding with the project description and objectives, it is interesting to indicate, once again, the two involved domains: GPS and its particular use for pedestrian navigation, which lead to the current research topic.

1.1.1 Global Positioning System

Nowadays, with the development of mobile devices and GPS technology more and more facilities like navigation systems, personal tracking, location based services, are accessible for everyone, wherever they are.

In fact, the position data itself is only helpful when it is combined with more information, like in the examples mentioned above. By combining the position of a person with a digital map or a geographical information system, the exploitation of the associated database provides indications on the direct surrounding of this person. Thanks to those developments, the civilian use of GPS keeps on growing and has been more recently followed by an increasing number of personal devices, mobile phones and Personal Digital Assistants (PDA) coupled with Bluetooth devices, offering navigation functions *(Steiniger and al., 2006).*

The best known GPS navigation function is the way finding application mostly used by car, truck or motorbike drivers. However, GPS technology is developing rapidly. More than « traditional » positioning and way finding, some online services are offered, continuously up-to-date according to the current location of the user. It calls the location based services (LBS), which expand to answer to the growing demand of ''Where's my nearest ...?'' *(D'Roza and al, 2003).* Indeed, an important development takes place in the domain of spatial data server. It aims to provide information about tourism, shops, traffic state, housing…By this extension of the available services, GPS technology opens up to a larger users panel and become attractive for a broad diversity of application, especially for walkers using pedestrian navigation.

1.1.2 Pedestrian navigation

At the moment, not only car drivers and motor bikers use navigation systems, but a larger public starts to be interested in those services. Hikers, climbers and even ordinary walkers in urban or rural areas use GPS to determine their position and orientation. Indeed, the development of GPS system is turning towards the pedestrian navigation. Based on the same principle as the one used for cars, it includes the wide range of applications. Currently this pedestrian navigation system, that visually supports orientation and way finding (*Baus and al., 2002, Gartner, 2004, Aslan and al., 2004*), is mainly used in urban area, limited by the available digital road network.

In the last few years, some recreational projects have been initiated in the field of tourism, like the Digitale Wichelroede project in the Netherlands (*Van Lammeren, 2006*), extending the use of pedestrian navigation with GPS to other networks. In those cases, the principle is based on GPS system sensing the tourist passing one of the points of interest in the database, and thus playing an audio story or giving information, and occasionally imploring directions. (*Gnatek, 2004*). Even considering this new ability, the GPS still keeps its basic function to offer positional data to guide people through an area. Moreover, the networks used for those projects remain few and limited to the application area.

Figure 1.2: A Location Based Service delivering a map of the environment and the position of the hiker (Steiniger and al., 2006).

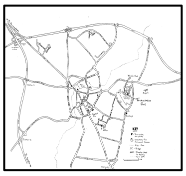
1.2 Hiking network

Now the foundation of the subject has been presented, some essential factors concerning the study components have to be set before to start dealing with the issue properly. This network is aimed to be used by walkers for a leisure hiking activity purpose. Thus, it generates some consequences about the data and the network, considering the target application and the final user. Some definitions are needed to what is called and meant by the terms user, hiking network, devices and data.

1.2.1 The user

In non urban landscape a wide range of possibilities are available for walkers to explore an area. The extent of those possibilities depends on the physical ability of the hiker and its behaviour. Indeed, some are more adventurous than others which prefer staying on easy and well indicated route. It leads to different ways to comprehend and approach the network. In this case, the largest user type is considered, in order to satisfy the majority of the users. It goes from really occasional walkers to most trained and experienced ones.

1.2.2 Hiking network's path



First of all, what can be considered a hiking path? This is an important question to answer before to go further into the research. A characteristic of such network mainly located in rural area is the possible changes occurring on small paths. Those latter can appear or disappear each year due to the natural and agricultural surroundings. Thus, some parts of a hiking network are too unpredictable and temporary to be taken into account. Only maintained and clearly defined tracks have to be included into the final network. A maintenance operation can be considered as sufficient when a check is done at least once by year. Moreover, a seasonality aspect may be added. Indeed, in the Netherlands some places are sensitive to flood then they are not practicable the all year, like some paths in mountain area that are not usable in the winter because of the snow. Then this seasonality aspect has to be taken into account when considering the hiking network. Only paths usable more than 9 months by year should be kept. The others could be integrated, but in a different way. Another point is that some paths require particular equipment, like climbing trails in mountainous areas. Those particular paths will not be presented in the hiking network either. The last consideration concerns the safety of the user. For example highways or too dangerous paths should not be included in the network.

According to these specific criteria, a definition of a path is formulated, which is the basic element of the hiking network:

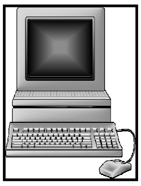
A hiking path is a track which is constant in time -because of maintenance action- can be walked without particular equipment -different than the one used for hiking- and without significant danger for the user.

1.2.3 GPS device

In this study, the terms GPS devices and GPS technology refer to commercial GPS equipments available to everyone in shops, about 285 euros, with prices starting around 150 euros (*[url16], 2007*).

Because of the increasing sales of GPS devices, many different models are available. If the basis function of satellite geolocalisation is the same for all, the main differences appear in the possibility to define the settings of tracking, precision and accuracy offered. Considering those latter, even if the capacity of the GPS devices differs, the current GPS horizontal accuracy is about 4 and 8 meters under good conditions (*Lahm, 2007, Doyle, 2005, Rupprecht, 2007*)

1.2.4 The data



As stated, this study focuses on digital data only and considers the existing data available via the Internet as well as the collection of new ones. A definition concerning the accuracy of those data is needed. This accuracy value will be used as a threshold value to check the relevance, the usability of the data and also to define the data collection methodology.

In comparison with the road network, a hiking network needs a better accuracy, since it is going to be used by walkers. For them a miss-guidance error in navigation of one kilometre has a more important consequence compared to a driver. Since no perfect accuracy can be reached and considering the characteristics of the project, it is about finding the sufficient accuracy? It leads to the question if a very high accuracy is needed. Of course, high accuracy level is always better, but it is often time and money consuming to achieve this. A good compromise has to be found. This case is dealing with GPS devices with a given current accuracy between 4 and 8 meters. Thus, it is not necessary to try to reach an accuracy lower than eight meters. Moreover, a navigation error, namely a misguidance leading to turn back, of at most 15 meters remains acceptable for hiking purpose. This value has been chosen because it is just less than the double of the maximum GPS inaccuracy (2 x 8 = 16 meters). It allows taking into account the GPS inaccuracy inherent to the network, if this one is collected with GPS devices, and the GPS that can use the hiker. The threshold inaccuracy value is set at 15 meters. However, the lower this value is, the better it is.

1.3 Research statement

The basis components of the project have been set and described. It is now about how the problem is tackled, which approach is developed and what the objectives of this work are.

1.3.1 Open mapping approach

In this thesis, the choice is made to treat the problem in a way that suits the future developments of open mapping. This approach has been selected, because it represents a real opportunity to create a good network, which is quickly and freely available to everyone. Moreover, the growing interest in open source projects (*Oreg and al., 2007*) and the fact that many people are already familiar with sharing hiking data, are two positive points supporting this idea.

If the problem is tackled according to the philosophy of open mapping, it will influence the way of thinking about the hiking network improvement. Then, the choice of the methodology may be a compromise between: moderate costs generated by the data collection and computation, sufficient accuracy and technical accessibility. Indeed, the methodology has to be applicable by everyone. The data are collected by volunteers, which are not specialists and don't have particular funding. Then the costs of the applied methodology, including equipment and software concerns, should be kept to a minimum. The methodology itself has also to stay technically simple to be usable by a large public.

As exposed earlier, a good accuracy is required for the creation of the hiking network, but a very high level of accuracy is not necessary. This fact allows to consider the open mapping concept approach by limiting the costs linked to the improvement of the accuracy.

In summary, the application of the methodology has to be kept cheap and accessible to a broad public which is not experienced, always respecting a relevant accuracy.

1.3.2 Research objectives

According to the described problem, the main goal of this thesis is:

"The improvement of hiking network for GPS users"

This thesis aims to establish a methodology to improve the hiking network for GPS users. This improvement concerns the digital hiking network, constituted of several data grouped in a database. This statement refers to three main topics:

- GPS users, which is the target group
- Hiking, which refers to corresponding network, route, track, etc…,
- Improvement, in other words, to see what is available and usable, and to complete it.

All of those consider a fourth component: the open source approach.

Based on this, two questions arise:

↳ **How to collect data?**

↳ **How to process data to obtain the final network?**

Concerning the first question, it will be carried out by exploring the available hiking data and collecting new ones in the field. The second question will be solved by processing the data to build a segmented network. This will be performed by keeping in mind the final user and the open source approach.

The first part of this study concerns the hiking data, in link with the first research question. It deals with the analysis of the existing data and the collection of new ones reminding the network with its particularities. Then, the first objective is to find, classify and evaluate the relevance and usability of the available hiking data to conclude on their quality and suitability for the purpose of the project. This will be developed in chapter 3. As not all the data are available, the second objective, developed in chapter 4, has to do with the establishment of a protocol to collect new data to improve the hiking network according to specified requirements. Chapter 4 is about the data collection methodology. Validation by means of field tests, it establishes the most suitable way to collect data.

- Find and classify the available hiking data (chapter 3)
- Assess available hiking data and conclude about their usability (chapter 3)
- Test different collection methods on field (chapter 4)
- Compare and assess different collected data to establish the collection methodology (chapter 4)

The second part of the project, corresponding to the second research question, concerns the data processing that leads to the improvement of the hiking network. The third objective is to compile the data into a network. After, the fourth objective has to do with the definition of useful attributes to describe and segment this network. Finally a field test will be carried out in order to validate the created hiking network. Both objectives are treated in chapter 5, which focuses on the network creation. It groups the data, the suitable existing ones, the collected ones and some additional ones, in a relevant network. It also deals with the network segmentation aspect.

- Set up a methodology to process the data into hiking network (chapter 5)
- Define attributes and segment the network (chapter 5)

The expected result for this research is the establishment of a protocol describing how to use, collect and process data to improve the hiking network for GPS users. Finally, this methodology will be applied for a practical example performed in a selected area. The validation enables to draw some conclusions, comments and recommendations, presented in chapter 6, for the use and a probable improvement of the protocol. In this last chapter, the methodology will also be discussed in relation to the French hiking network created by the IGN (National Geographic Institute).

In order to fulfil the research objectives outlined above, a step-by-step approach is performed according to the methodology developed in the chapter 2. This latter offers the main methodology which is used to carry out this project and the principle analysis applied to the data.

II. Methodology

This chapter offers a guideline to the methodology followed during this thesis work. After having stated the methodology approach, according to the requirements, the study area and the field tests performed are presented. Then the different analyses carried out to assess the data and network are described.

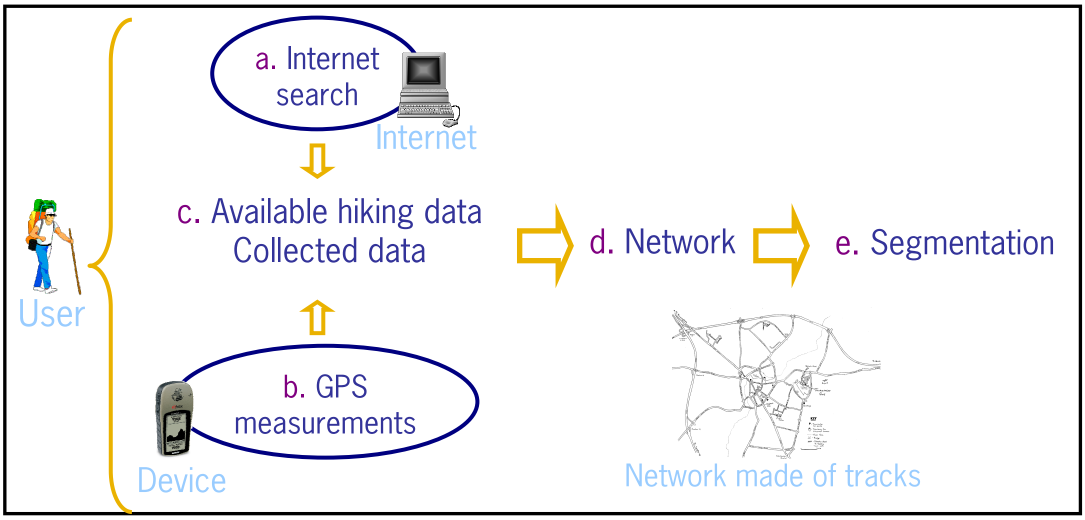


Figure 2.1: Followed methodology.

Figure 2.1 shows the idea to go from several sources of data to a coherent segmented network. To do this, a four steps methodology is followed. First the hiking data available via the Internet is explored and assessed [a]. Second, new data is collected [b]. Available Internet discovered data and the collected are compared and combined [c]. Based on the findings, the network is constructed [d] and segmented [e].

2.1 Methodology approach

Based on the open source approach, it is possible to make a first selection from all the existing data collection methodologies exposed in the literature. The objective is to get data easily and cheap. Thus, using already existing data before to collect new ones is also a possibility to explore. After, the creation of a network and possibilities for segmentation are discussed. The objective is to create a suitable network for walkers from the available data.

2.1.1 Hiking data

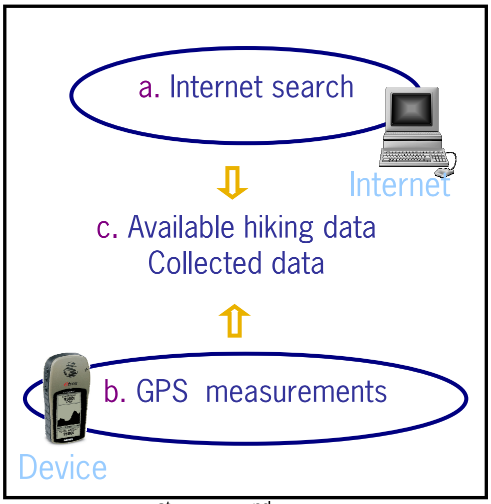


Figure 2.2: 1st and 2nd steps of the methodology.

As shown in figure 2.2, which represents the first three steps of the methodology, there are two ways to acquire data: exploring the existing ones and collecting new ones.

It is known that a lot of data are already available via the Internet. Then those data need to be found and assessed to be found suitable or not for the purpose before being used.

The second way quoted to acquire data is to collect them. The point here is to find the best collection technique, based on literature, to fit with the open source approach requirements. It is carried out by considering the feasibility and the limitations of all the possible methodologies, but also the particularities of the hiking network mainly located in rural area and destined to walkers.

Different documents already described some applied methodologies to update road network data to the benefit of hiking network. In their paper *C.A. Quiroga and D. Bullock (1998)* illustrate the possibilities of the GPS/GIS methodology for data collection procedures. Recent developments in the field of digital tracking technologies have generated a range of widely available systems including land-based tracking, satellite navigation, and hybrid systems. The study, carried out by *Noam Shoval and Michal Isaacson (2007)*, summarizes and experiments those methods to collect data to the extent of spatial and temporal activities of tourists.

The selection is based on cost and accessibility limitations. Then, all methods using complex and expensive equipment, can be set aside of the project. So, DGPS or RTK devices, as well as inertial navigation system and radio frequency are considered as non-suitable methods. Also, photogrammetry is not relevant, mainly because this method may need some software and knowledge, which is above the broad user accessibility requirements. It is possible to find free images or to use Google Earth to solve the problem of the costs. Digitalisation uncertainty, which is inherent to this method due to the base map used, the digitalisation work and the projection error that can be done, remains. In addition to the cost concern, this method is not reliable enough, due to the lack of information.

Based on this, two possible methods remain: GPS measurements or land based antennas network, like cellular triangulation. According to field tests, this latter turned out to be suitable in build-up area, less subject to signal degradation and lost of signal than GPS, but restrained by a signal reception setting fixed and broad, which can affect the accuracy (*Shoval and al., 2007*). Moreover, this method doesn't provide a suitable accuracy in rural area, where the data collection methodology will be applied mostly. Thus this method can't be used to reach the project objective, because of accuracy limitations.

Concluding, the most suitable option is GPS technique. According to the cost concern, those materials become more and more cheap and available for everyone.

Furthermore, they are easy to use and don’t ask for any particular knowledge or training. The most critical point is the accuracy. Nowadays, there are good devices accurate enough for the objective. However, there are some conditions, like under vegetation cover, indoor measurements or in build-up area, where the signal is degraded. This results in a decreasing accuracy of the measurements. This is one of the main drawbacks of the GPS technique. However, this limitation doesn’t make this method unusable here. Indeed, the indoor limitations are not a problem in this case, because all the measurements are taken outside. Concerning the weak signal in urban area, the problem is solved by the existence of the necessary information. What remains is the signal quality degradation under the vegetation cover. However, since few years, the precision under forest improved and will go on this way (*Yoshimura and al., 2005*). Furthermore, when Galileo will be available this problem will tend to reduce a lot (*Weimann and al., 2006*). Some tests performed during the integration course in the Achterhoek, shown that the accuracy decreases under the trees cover, but still remains acceptable for navigation purposes (*Bulsink, 2007*).

Before going further, some terms need to be defined. The report deals with the GPS technique, which causes the use of specific vocabulary. Three types of data are related to GPS: waypoints, tracks and routes.

➥ A “**waypoint**“ is a reference point in physical space used for navigation purposes.

➥ A “**track**“ is an ordered set of points along a path, collected automatically by the device and represented as a line data.

➥ A “**route**” is defined as a series of one or more waypoints in GPS navigation. A route is followed by navigating to the nearest waypoint, then to the next one in turn until the destination is reached.

2.1.2 Network and segmentation

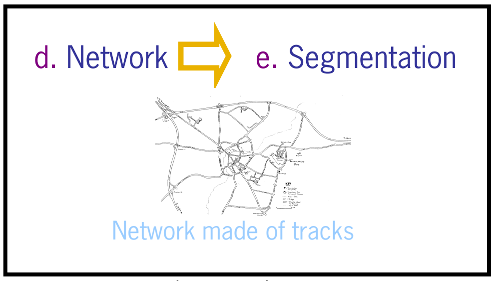


Figure 2.3: 2nd and 3rd steps of the methodology.

The last two steps of the methodology deal with the construction and the segmentation of the network.

The establishment of a hiking network differs in several aspects, due to the pedestrian transport mode and its particularities, from the already well-known road network. This network represents its own characteristics. “If the principle of navigation remains the same, the models used are, them, completely different” (*Gillieron and al., 2002*).

The data can be processed in order to get a suitable hiking network. However, the principles used for road network processing can not be directly applied to pedestrian network, because preconditions differ significantly. While vehicles can only move along predefined paths, pedestrians have a very high degree of freedom in exploring their environment. There are many different possibilities of how a square can be crossed and barriers, like greens and fields, can be passed without following a certain path (*Gartner and al., 2005*). Some existing methodologies are presented in some tourists’ behaviour studies, where data has been collected and processed to obtain the corresponding trip. In their paper, *Y. Asakura and T. Iryo* (*2007*) propose a protocol to process tracking data collected with a mobile communication instrument.

Also A. O'Connor, A. Zerger and B. Itami (*2005*), offer a methodology via their geo-temporal tracking and analysis of tourist movement. Those studies provide indications about how to process data for this project purpose.

2.2 Study area and field tests

In order to test the existing data, to establish the most suitable data collection methodology and to validate the final result, some field tests are carried out through a concrete and practical application in a selected study area.

2.2.1 Study area

The study area is located in the Netherlands, between Ede and Apeldoorn, in the Veluwe area (Figure 2.4 and Appendix1). Because of the presence of the national park, this area offers many hiking paths and already available data, interesting for the current project. Indeed, it allows to assess the available data, to test different collection methodologies and find the most relevant.

Figure 2.4: The study area, Nationaal Park de Hoge Veluwe, between Ede and Apeldoorn, in the province of Gelderland.

2.2.2. Field test

➯ **Assessment of the available data on the field**

The available hiking data offered via the Internet are evaluated in two ways. First, theoretically according to criteria defined in the next paragraph (data analysis). Second, by using the data to navigate in the field, which refers to network assessment defined later on. This has to do with the compatibility of the data by integrating them into a network.

Test different collection methodology

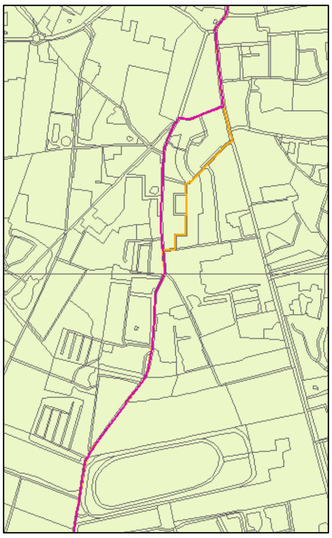


Figure 2.5: Fieldwork walking path.

GPS technology offers several ways to collect data. Indeed, it can be done automatically by using the tracking function or manually by taking waypoints all along the trip. It is about finding the most suitable method to collect walking path data and the most relevant setting for this technique. To do this, field tests are carried out with two different devices, a Garmin and a PDA combined with a Bluetooth GPS receiver. Tracks and waypoints data are collected by different users and according different settings.

To collect the experimental tracks, a trip is designed through the study area. It is designed to take into account different surroundings, like build-up area, open spaces and forest. Almost the same way is travelled on the way to go and the way back. On figure 2.5, the way to go is represented by a yellow track and the way back by the purple one.

On the way to go, two persons collected waypoints with two similar Garmin devices. These two users didn't receive any indications, in order to get information about the intrinsic personal influence on the way of collecting waypoints. At the same time, a third person was guiding the group through the area helped by the network coming from the existing data. The two PDA and Bluetooth GPS were tracking the trip with different collection settings, one based on time (1 point every second) and one based on distance (1 point every 3 meters). The Garmin used to collect the waypoints were also tracking according to its inherent collection setting which is not adjustable.

On the way back, the person who didn't collect waypoints previously did it. Another person was guiding and the last one was taking waypoints only when a turn was occurring in the followed route. The PDA and Garmin were still actively registering the track. For the way back the tracking setting of the PDA have been changed, but still one based on time (1 point every 5 minutes), and one based on distance (one point every 50 meters). The Garmin devices were still collecting with automatic settings. Figure 2.6 offers a scheme of what was achieved during the field work and table 2.1 summarizes the tests performed on the field.

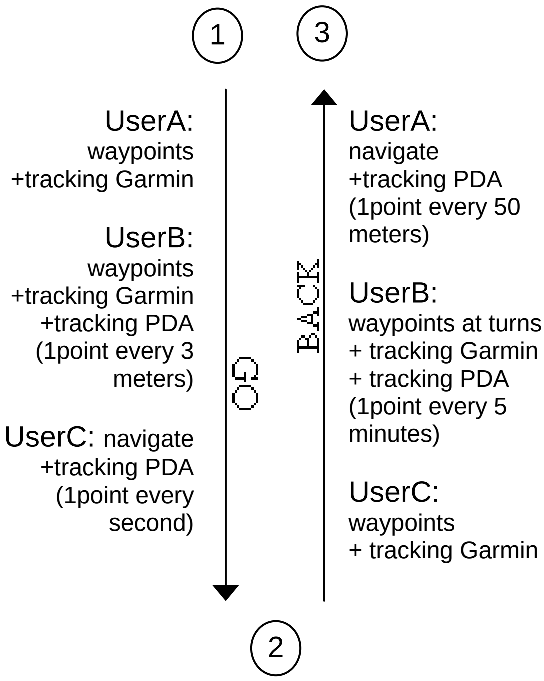


Figure 2.6: Data collection scheme.

Devices	Collection	Settings used	Way
Garmin	waypoints	UserA	go
		UserB	go
		UserC	back
		Turns	back
	track	Automatic gps1	go
			back
		Automatic gps2	go
			back
PDA+ Bluetooth	track	1 sec	go
		50m	back
		3m	go
		5 min	back

Table 2.1: Fieldwork data collection plan.

The idea in this methodology is to check different elements:

- the influence of the method by using waypoints or tracking collection methods,
- the influence of the devices by using two types of devices,
- the influence of the settings applied to the data collection (waypoints regularly or only when turns occurred, tracking based on different times or distances)
- and the personal influence of the user.

➭ **Collect data according to this methodology**

When the acquisition methodology was established, testers went on the field to collect data according to the requirement. Few data were collected because the objective was to use them to construct the network. Then it was more of interest to send people collecting several time the same information in order to check the user influence; instead of collecting a lot of data when the result expected is not a complete network, but a relevant protocol. Then three testers collected some paths according the defined methodology presented in chapter 4, in the study area.

➭ **Validate the network on the field.**

The best way to evaluate the sufficiency of a network is to use it to navigate in the field. It allows to evaluate the sufficiency of the network by checking if the decisions and indications brought to the user are suitable for hiking navigation. It has been carried out during the fieldwork. During the assessment, the people were following a computed route displayed on the PDA with the road network as an underlying layer. This route has been chosen to cross different possible surroundings, to take into account the fact that the GPS signal could degrade under vegetation cover and build-up area. The navigation test is realized by different persons in order to not be biased but individual characteristics. The last requirement will be that no one of the user will know the test area in advance. The result is made of personal impressions of the testers, but also by comparing the computed route with the travelled one, which was collected during the test.

2.3 Data analysis

2.3.1 Existing data

It is possible to find digital data about walkways, there is a lot available via the Internet. An important question to answer before using those data for the network construction concerns their relevance and their accuracy. Thus, criteria are required to assess them. The first point is to define relevant criteria by wondering what is of interest about those data. The two main questions are: can I use the data? And how can I use them? Those questions refer to the availability and the accessibility aspects. It concerns the reliability of the data according to their inherent information and characteristics.

The availability deals with the indication of the potential usability for the network processing. It concerns all that brings information about the data themselves, which can be described in detail by different points:

- **Metadata quality**. This refers to the estimation of the offered data description. Is it complete and relevant?
- **Data accuracy**. It deals with the possibility of knowing the data accuracy and defining if it is sufficient for the purpose?
- **Collection method**. This criterion focuses on how the data have been created, with which means and what is the type of information collected (track, route, and waypoints).
- **Updating**. It relates to the relevance of the data according their creation time, but also if they have been modified or updated after all.
- **Data quantity**. It focuses on the effective proportion of walking path data available.

On the other side, the accessibility criterion focuses on the legal part of the data. It concerns aspects dealing with how to acquire the data and the right to use them, which is defined by the following points:

- **Authorization**. It deals with the rights and the restraints linked to the use of the data.
- **Costs**. It relates to the possible costs to obtain the data. There are three possibilities: the data can be provided freely, against fees or can be sold.
- **Format**. It concerns the possible formats in which the data are provided.

Those criteria haven't been chosen randomly, they come from a specific selection. According to the ISO19115 document, the reference for the spatial data metadata standard, many attributes and information should describe the data. The criteria mentioned above have been chosen amongst others, because of their relevance to the subject dealt with.

2.3.2 Data accuracy assessment

To assess the available and collected data, different factors can be calculated and analysed. It allows data comparisons and visualization of the interactions between the different factors.

A. Observation

The first analysis to be carried out is a simple observation of the tracks by displaying them at the top of some reference data as top10, top25, which are the Dutch Topographic datasets. Indeed, just by looking tracks collected with different settings, it is possible to draw some conclusions, like a first classification of the accuracy and the smoothness of the track.

B. Mean deviation

Besides the observations, the mean deviation between the data and a reference with a known accuracy can be calculated. In that case the term deviation refers to the distance in meters between the theoretical and the collected location of the track. That means the distance between the collected track and the referenced one.

The top10 vector dataset is used as reference data (1). From this, the theoretical travelled path is digitalized (2) by drawing it in the middle of the followed path. The collected path is displayed on the top of it (3). The polygons representing the area between the data and the reference are digitalised (4). From the data statistics of the polygon layer it is possible to know the total deviation area, which is the sum of all the polygons areas (5). Then, this value is divided by the length of the collected track in order to calculate the deviation mean, expressed in meters.

Figure 2.7: Mean deviation computation.

Figure 2.7 shows the collected path (in red), the reference track (in purple) displayed on top10 data, and the deviation polygons (in green).

In addition, the location of the deviation polygons can also be checked to learn more about possible external influences. Other factors, like the **number of polygons** or the **mean area of those polygons**, also bring relevant information concerning the assessment of the data. Those values are also provided in the statistics summary of the polygon layer.

The polygon deviation method is applied to compare tracks with a reference, but it can also be used to compare the different tracks.

C. Point density

The point density value refers to the number of points for a given distance. It is calculated by dividing the length of the track by the number of points collected. This value can be directly linked to the data accuracy. It also gives an indication about the potential quality of existing data.

The point density value is given as the number of points for 100 meters. This factor is based on distance instead of time, because the length of a track is a known and secure value whereas the travelling time is an estimation. Indeed, the distance value is obtained by measuring the track and the time value by calculations according to the potential walk speed of the hiker. As the walk speed can't be known accurately, the result would be less precise.

D. Smoothness

The smoothness of a track is evaluated by a visual assessment. A value, between 1 and 5, is attributed to each track. The value 1 corresponds to a very fuzzy track and the value 5 to a clear and smooth one (Figure 2.8). During this evaluation, the accuracy is not tacking into account, only the aspect of the track.

Figure 2.8: Smoothness scale.

The different factors used to assess the data are summarized below:

- **Observation (A)** of the data displayed at the top of reference data.
- **Mean deviation (B)** value. It is the average value of the difference between the collected and the reference top10 data.
- **Number of deviation polygons (B1)**. It corresponds to the number of polygons obtained from the deviation calculation performed in figure 2.5.
- **Mean area of the deviation polygons (B2)**. It is the average of all the polygons area obtained from the deviation calculation performed in figure 2.5.
- **Point density (C)**, which is the number of points collected for 100 meters.
- **Smoothness (D)**, which gives a value to the track considering its smoothness.

A Pearson correlation analysis can be performed with SPSS, to see the interactions between those different factors and to define the influence of each on the track accuracy.

III. Available Hiking route data

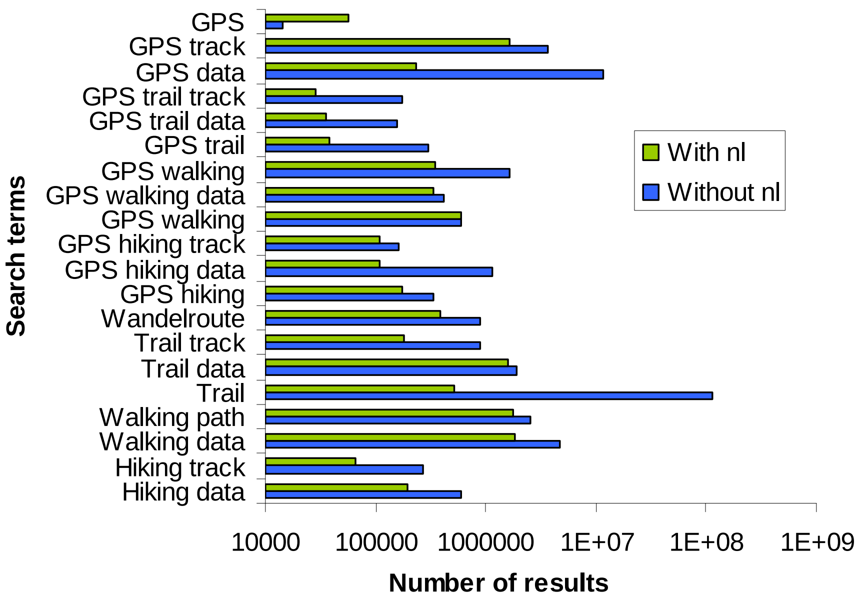
Nowadays, a large quantity and diversity of digital hiking data are available. The objective of this part is to find, describe and assess those existing data to conclude about their usability in the hiking network. To do this, a research procedure is applied in order to get the best overview of what is currently offered via Internet. Thus, the findings are described through an established typology. After that, each data type is assessed according to the availability and accessibility criteria described previously in the methodology. The compatibility of the data is checked by integrating them in a network. Finally, this evaluation concludes on the selection of suitable data.

3.1 Available data overview

3.1.1 Web search

The starting point to analyse the available data, is to get a representative view of what is offered. To do this, a search via the Internet is performed using different keywords: "track, trail, data, hiking, walking, wandelroutes, Netherlands, GPS" used alone or combined. For each search the number of results is checked as well as the relevance. This latter is determined by the number of websites providing hiking information in the first two pages of results. A distinction is made according to the information provided by the site. It can be GPS data, other kinds of data or links to other interesting websites. This experiment is carried out with Google search engine using the standard mode (the first findings come from a search with "AND" as Boolean term and the next ones from a search with "OR" as Boolean term). The search is kept simple to be as close as possible from what could be done by a user.

Figure 3.1 presents the hits obtained for each keyword, or keywords combination.

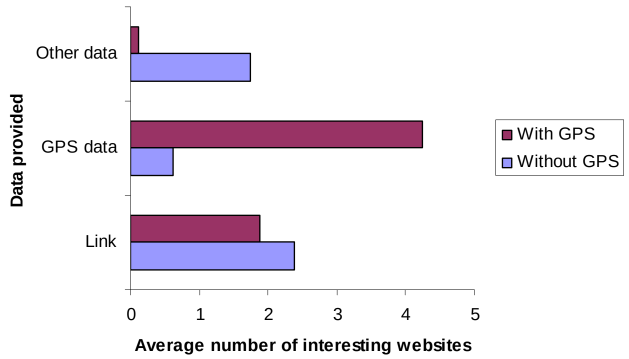


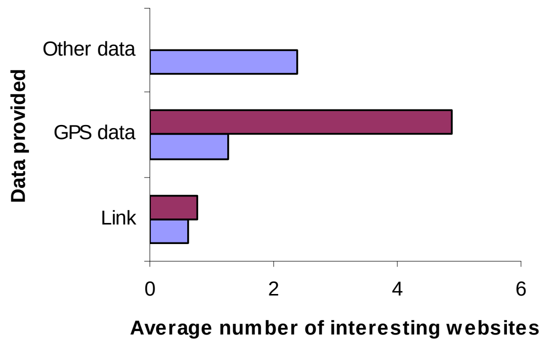
Each search is run twice: one with the use of "Netherlands" as additional keyword and one without.

For "wandelroutes" and "GPS wandelroutes" terms, the use of "Netherlands" has been commuted by "Nederland".

Figure 3.1: Number of results according to search terms.

To go further, a comparison is made to see the influence of the keyword "GPS" on the kind of information provided by the websites. This test is also performed with a difference made between using (figure 3.2a) or not (figure 3.2b) the term "Netherlands". Figures 3.2 show the average number of websites offering interesting hiking information, according the kind of data provided.





Figures 3.2a and 3.2b: Search hits by data type with (a) and without (b) "Netherlands" term.

Figures 3.2 highlight the direct influence of the term "GPS" on the type of data found. There are about 3 times more relevant results for "GPS data" category and 2 times less for "other data" category. However, there is no significant influence on the websites providing links. The use of "GPS" in the Internet search allows to find more relevant websites and to focus the findings on GPS data.

3.1.2 Data typology

After having checked the majority of the data offered via Internet, let's explore the different data found. There are several ways to describe and classify those data. An abundant literature discusses about digital data typologies (*Adrienko and al., 2007)*. They are based on storage, data management or focused on the nature and inherent properties of the data. However, they mainly concern spatio-temporal data. According to the current scope dealing with a more global scale (digital data), those typologies are too restrictive. Moreover, the interest of the data classification is also different. The objective is to evaluate the availability and accessibility of the data. It requires an adapted typology is required. The data type approach is chosen to tackle this problem. The different data types defined are:

- **GPS data**. This category refers to all the data collected with GPS-devices. According to the websites, there are provided as files containing waypoints, tracks or routes. However all the data provided are not available in all the formats (track, waypoints and route).

- **Digital Maps**. It concerns all digitalized maps available in the "old fashioned" way, by means of hard-copy. They may come from booklets content, or other sources that offers picture like Google Earth, put in a digital form by scan or redraw. There is a broad variety of those maps. They can be more or less detailed, with the presence or not of complementary information.

Figure 3.3: Example of digital map [url15].

- **Geo-data or geographic data**. It refers to geo-referenced data made of features described by attributes. They are digital data with a spatial dimension which aim to represent element of the real world.

The distinction between those data types is based on the level of information they bring to the user beside the path information proper. The digital maps bring only an “image of the path” almost never offers geo-referenced, that can’t be used directly in a GPS to navigate. The GPS data are the raw geo-referenced data produced by devices, sometimes completed with few attributes. The geo-data are the most complete; they are mostly geo-referenced, precise and well-described.

Afterwards, each data type can be detailed according the provider information. The following main categories raised during the data researches:

- ***Official data*** are the data produced and provided by official society, governmental, like the Kadaster in the Netherlands, or private, like Teleatlas which is one of the main provider of geo-data for GPS.

- ***Organisation data*** are the data made available to all by organisations, for example tourism centre or hiking club.

- ***Personal data*** are data collected by individuals as leisure or in order to share experiences.

It is possible to relate data types classification to the provider information. However, no precise percentage can be presented because not all the available hiking data found have been checked. Even if the number of sources consulted was broad, it can’t concern all available data. Nevertheless, it is possible to formulate some comments about the relation between data types and providers.

Digital maps and GPS data are unequally spread between organisation and personal data. The official data are mainly geo-data and occasionally digital maps. It is important to specify that on Internet websites, the distinction between organisational and personal datasets is not that sharp. Indeed, GPS data are divided between the organisation and the personal provider categories defined previously. However, in reality, most of the personal tracks are offered more via websites of organisations, than via personal web pages.

Another important point is related to the total amount of geo-data, mostly offered by authorized institutes and companies. Since few years some geo-data are also provided by open source projects. It still remained a small proportion of the geo-data. However it keeps growing and tends to gain more and more importance according to the increasing interest in open source software and open source initiatives observed during the last few years (*Oreg and Nov, 2007*).

3.2 Data assessment

3.2.1 Data availability

GPS data:

Concerning GPS data, there is almost no metadata available. The only information offered to the user is a short text description of the walk itself, with helpful indications like the location of the starting point, the length, the duration of the trip... Some more practical information may be added, as if there are restrictions concerning animals or ticket entrance for example. There are also complements about the difficulties or the interesting particularities of the route. Sometimes pictures are added to complete the description. Some websites (2 out of 15 websites during the web search) allow other users to add their personal opinions, comments and advices about the trip. Even if those area descriptions are rich and complete, they don't bring any interesting information about the data themselves and how the data have been collected.

The main consequence of this lack of formal description is that it is difficult to evaluate the data accuracy. However, as those data have been collected with GPS device, their accuracy can be assumed equal to the one of the device. Then, the current knowledge about this technology allows to get an idea about the data accuracy and the deviation of this accuracy (*Naus, 2006*). Moreover, according to experiments carried out in Achterhoek, GPS devices appeared to be suitable to track and navigate walkers on the field (van Rooij, 2007; Gijzen, 2007). As proved by those studies, the accuracy of the GPS data is sufficient for the navigation purpose, in spite some negligible misguidance occurring sometimes.

The collection method used is never described either. But, according to the datasets provided (track, waypoints or route), this information can be identified.

The GPS data offered are not often updated whereas new information is added regularly. It is sometimes possible to find the same track in a more recent version on other websites.

In summary, there is a lot of GPS data available via Internet. They are provided in different forms and from different sources. Those data are spread and there is currently no document, principles or directives, applied to harmonize and standardize them in such way that they could easily be grouped into a suitable hiking network. Even if the quantity and importance of GPS data increase, it is still difficult to get precise information about their accuracy. However, it is possible to have an idea about it thank to the knowledge concerning this technology and the data collection method.

Digital map:

As GPS data, they often come with a complete trip description but almost no metadata. It provides to the walker necessary information to travel safely, as well as indications about the interesting characteristics of the walk and surrounding. But few information concerning the use of the data is known. The data used as basis to digitalise is almost never specified. Thus, the accuracy is highly uncertain. There is no indication about how the walk has been digitalised either.

In opposition, to GPS data for which the accuracy can be guessed, any information can be deduced from digital maps if there is no indication given about the base map and about how they have been created. Furthermore, those data are rarely updated. Then, the uncertainty of those data is too important to be negligible. An experiment shows that digital maps created from top10 data or Google Earth are suitable to digitalise walk trip (*Roupioz, 2007*). But because of the uncertainty reasons explained before this king of data won't be kept for this work.

Digital maps represent a large amount of work to adapt them, when it is possible, to this project requirement. Moreover, those data start to be replaced by GPS data. Few years ago, digital maps and paper maps were the main information available about walking paths. Nowadays they tend to disappear in favour of GPS data.

Geo-data:

Concerning geo-data, the metadata, the accuracy and the updating are always known and satisfying. So they can be used securely. However those data represent only a small proportion of walking paths, that are of interest in this case, and the information is mostly located in urban area.

On the website of Mappy ([url17]), which is an itineraries' provider based on Teleatlas information, it is possible to see that some pedestrian paths are available in urban area, and that they are taken into account during the computation of pedestrian trips.

Then, according the availability criteria, those data fit to the criteria even if they represent only of small part of the real network.

Data Category	Metadata provided	Accuracy	Collection method	Update	Quantity of walking paths represented
GPS data	Rare	According device accuracy	Waypoints or tracking	New data often added, few update	Many data available
Digital map	None	According source data, digitalisation quality	Draw by hand or scanned from different data sources	Few update	Many data available
Geo data	Complete	Different accuracy available	From paper map, field measurements, photogrammetry	Well update	Lot of data in urban area, few in rural area

Table 3.1: Data availability criteria summary.

Table 3.1, summarizes the availability data assessment for each data type according to the different criteria. From this availability analysis, the GPS and Geo-data appear to be suitable for the purpose. Even if GPS data are poorly described, it is possible to get a sufficient idea about it.

3.2.2 Data accessibility

GPS data:

According to the website, different downloading formats and displaying facilities are provided to users. The most common export format is ".GPX", which can be used for waypoints, tracks, and routes. ".TRK" and ".WPT" formats, which respectively concern only tracks and waypoints data, are also offered, as many other formats. Sometimes the data can be directly exported to a Garmin device. One of the visited website was offering the data as coordinate in a table in text format. This is the less handy possibility but remains one case among all the websites checked.

Most of the websites offer the possibility to display the datasets via KML or GPX format on Google map and Google earth. Doing so allows users to visualize clearly the trail and also to modify it.

The access to those data is mainly free and easy, even if some websites require registration or some fees before to get it.

Digital map:

They can mainly be downloaded in image format from organisation websites but also from personal pages, like blogs. Those data are mainly free to use and without restrictions. However, some organisations can ask for fees. According to the findings of this study it represents around 30% of the cases. Digital maps can also be images scanned from a book or from official paper maps that can be used as base map. In this case, their use is often under restraint.

Geo-data:

Those data are subject to serious restrictions and are most of the time really expensive, when they are provides by governmental or private companies. For many years, those latter were the only providers. However, as quoted before, some open mapping initiatives upraised to create and share those geo-data through open source application based on a wiki principle. Then they are free and without use restriction.

The accessibility is mainly linked to the notion of copyright of data and the cost, summarized in table 3.2. The format criterion has been discussed above.

	Copyright	Costs		
Group		Free	Participation fees	Full cost
Official data	Authored			Geodata
	Non-authored			
Organisation data	Authored	Digital maps GPS data	Digital maps GPS data	
	Non-authored	Geodata Digital maps GPS data		
Personal data	Authored			
	Non-authored	Digital maps GPS data		

Table 3.2: Data availability criteria summary.

From the accessibility analysis, only the GPS data fulfil the requirements of the project: cheap and accessible. Indeed, the digital maps are not suitable because of the format in which there are provided and the geo-data because of the costs and restrictions. Even if it is possible to find some free geo-data provided by open mapping initiatives, they unfortunately offer almost no information about walking paths. Then, they are not on interest here.

3.2.3 Assessment conclusion

The previous availability and accessibility analysis leads to the conclusion that the GPS data are the most suitable data to use for the purpose of this study; because they are most of the time free or low cost data and without restrictions. Indeed, GPS data are preferred to the digital maps which are too unpredictable and require too much work. The geo-data, even if they are the most secure one, are too expensive and restricted, beside the fact that they bring only few interesting information about walking paths. Moreover, many GPS data are already offered and the amount of GPS data keeps growing. They would be very useful for the construction of the network and it will save money and time. However, because of their diversity and the lack of information, the data have to be used carefully. They have to be homogenized and tested to be considered as suitable to be integrated into the network. From this existing data analysis, the main need raised was the need of information model: a formal description of data.

3.3 Data compatibility

According to the availability and the accessibility criteria, the GPS data appeared to be the most suitable data. However, the problem concerning those data is their broad heterogeneity. They are spread between different websites. They don't have the same accuracy and characteristics because of the use of different possible devices and collection method. Moreover, as underlined previously, they are poorly described. Then, those GPS data have to be checked to see if they can be used together. It refers to the compatibility assessment. To do this, a sample of GPS data is selected via the Internet. Then, they are homogenized and grouped together in a network (Figure 3.4). Afterward the result is evaluated to conclude on the compatibility of the data. The process applied to compute the data into a network is the processing protocol developed in chapter 5.

3.3.1 Data selection

As there are too many websites to evaluate and compare all of them, only the most representative ones are kept for the analysis. A website is considered as representative when it is currently maintained and updated by an administrator. It also has to be visited by users regularly, the last visit should not be older than one week. As the study area is defined between Apeldoorn and Ede, the website may offer data for this area. According to those criteria, four websites are selected offering five different interesting datasets.

In order to represent the heterogeneity observed in the reality, it is important to respect the diversity of the data. To do this, all the data used come from different websites and are downloaded in different formats. As quoted, the GPS data can be provided as waypoints, tracks or routes. The main format observed is the track in a GPX file. The case dealing with websites offering only waypoints data has been seen one time and does not provide data for the study area. So the sample for this study is made of four tracks and one routes datasets, but no waypoints because they are less provided and not available for the area. One of the tracks is also available as route dataset, it will allow comparison. Then, two routes are used.

From those criteria, four websites have been selected that offer five different GPS data around Otterlo:

- **Ardoer**, a route file coming from www.gpsies.com
- **Bennekom**, a track file from www.gpstrack.nl
- **Ede-apeldoorn**, a track file from www.gps-info.nl
- **Ede-Harskamp**, a track file from www.gps-info.nl
- **Veluwe**, the same path in route and track files from www.ontrack.nl

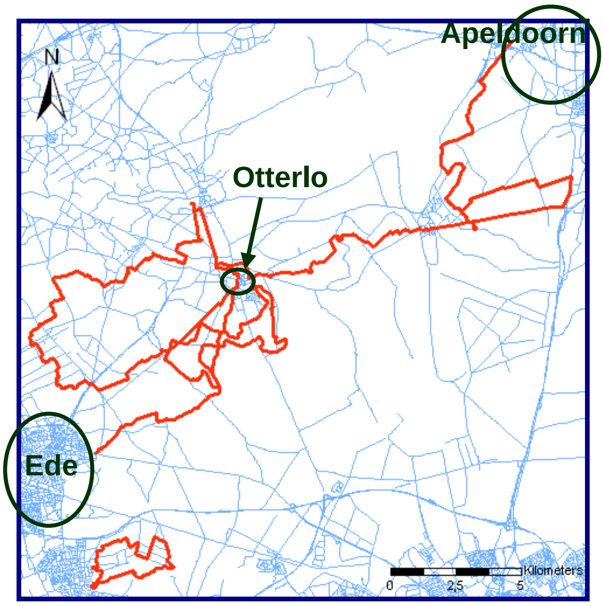


Figure 3.4: Network of the selected existing data.

3.3.2 Compatibility assessment result

The compatibility of the data is assessed in order to know if the data, even if they fulfil the accessibility and availability requirements, are usable together in a network. Indeed, the two previous assessments were related to the individual characteristics of the data whereas compatibility deals with the comparison and the gathering of the different data according to each other. It is performed by testing and evaluating the network created out of the data. To do this, some criteria defined in the methodology part to analyse the data and assess a network are applied.

• **Navigation test**

Even if a misguidance occurred in the city of Otterlo, the network appeared to be sufficient enough to navigate users in the real world. The high paths density in the residential area can explain this misguidance. It would have been easy to get back on the good track by turning back, but it was an opportunity to use the network to get back on the good track by another way, as shown in figure 3.5. The followed network is represented in red and the walked way in purple.

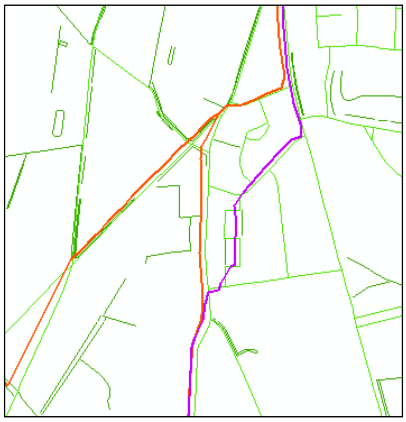


Figure 3.5: Misguidance in Otterlo.

The accuracy of the existing data used to create the network seems good enough for a walking navigation application, it allows to navigate on a predefine route and to find back the initial route in case of misguidance. This conclusion is also true in the forest, even if sometimes, when there are two very closed paths at a junction, you have to get a bit in one of the path to see if it's the good one. But this is also due to the inaccuracy of the device used to navigate, and not only because of the network. However with a general overview of the network, mistakes become less important that the track information alone is available. The surrounding information brought by displaying the full network helps the navigation.

• **Point density**

This factor allows to evaluate the difference between route or tracks datasets, and the internal variability existing for the same format.

		Points	Length (m)	Points/100m
Track	Bennekom	95	11759	0,808
	Ede-Harskamp	799	30721	2,601
	Ede-Apeldoorn	923	42565	2,168
	Veluwe track	305	20313	1,501
	Average:	530,5	26339,5	1,770
Route	Veluwe route	64	19056	0,336
	Ardoer	34	10486	0,324
	Average:	49	14771	0,330

Table 3.3: Existing data point density.

From table 3.3, the point density for tracks datasets is around 1,8 points per 100 meters. This value is much lower for the route, 0,33 points per 100 meters. So the accuracy of the route data can be assumed as lower than the track one. The point density values for track data are more variable whereas it seems more homogeny for route datasets. But only two route datasets are analysed here. According to the conclusive network test described above, a point density of approximately 1,8 points per 100 meters can be assumed as sufficient for a track. This value can be used as an indicator to get an idea about the usability of a track.

• Mean deviation

The methodology developed in chapter 2 to evaluate the data by calculating the mean deviation value, is performed on the existing data:

		Length (m)	deviation area (m^2)	mean deviation (m)
Tracks	Bennekom	11759,12	53568,88	4,56
	Ede-Harskamp	30721,55	140549,85	4,57
	Ede-Apeldoorn	42565,94	152550,09	3,58
	Veluwe track	20313,89	96357,18	4,74
	Average:			4,36
Routes	Ardoer	10486,11	31978,01	3,05
	Veluwe route	19056,56	210775,07	11,06
	Average:			7,06

Table 3.4: Offered GPS data mean deviation.

The average mean deviation obtained for all GPS data is 5,26 meters. If the data format is considered, tracks data present a deviation error of 4,36 meters and route data a deviation of 7,06 meters. Whatever the format, the calculated error is under the acceptable threshold value of 15 meters, set in the chapter 1. Moreover the location of the deviation polygons shows that the deviation is located most of the time at punctual position. It means that the errors are probably due to temporary decreases of the signal quality or the effect of the surrounding.

This analysis points out a difference between paths collected with the tracking mode of the GPS and routes files issued of waypoints. Even if both deviation values remain under the threshold value of 15 meters, it can be conclude that track data are more accurate than route data. But this affirmation has to be moderate by the fact that few route files have been tested. The thing that turns out is that, at the contrary of point density value, the route data present more variability than track data. Thus, route data are more uncertain to use. It is preferable to use track format data and fortunately it is the most available one.

This compatibility assessment shows that GPS data give a relevant network for the hiker navigation. It also demonstrates that some GPS data are more suitable than other. Indeed, the quality difference between routes and tracks is noticeable. From those tests, performed in order to represent most of the GPS offered data, it is possible to say that they can be used for the purpose of this thesis. The point density value can be used as an indicator to ensure the choice when data are finding.

Then, the GPS existing data are a good base for the improvement of the hiking network. It saves time and money from collecting already existing and usable data and gives the starting point to collect the missing one.

3.4 Conclusion

In this part, available hiking data have been classified according three data types: GPS data, digital maps and geodata. Those data types have also been approached regarding three provider categories: official, organisational and personal data.

It appears that among all available data offered via the Internet, some can be used to construct the hiking network. Between the three data types, only the GPS data appears to satisfy the availability, accessibility and compatibility requirements and especially the GPS track data. However, a lack of standardization and a broad heterogeneity have been pointed out. Then, those data will most probably require some adaptations before to be compiled in the network. The lack of metadata and, more generally, information about the data is an important issue here and a formal description has to be established.

The main conclusion about the data assessment is that there are many GPS datasets available for walkers who want to hike with the help of GPS devices but it represents only a small part of the total amount of walking tracks in reality. Moreover not all the existing data can be used to construct a network. This means that new and additional data has to be collected to extend and improve the network.

IV. Data collection methodology

As stated, a large amount of hiking data is offered via the Internet, but only GPS data are relevant for this project. Even if, there are more and more of those data, it still represents a small part of the real world network. Then, new data has to be collected. This chapter deals with the establishment of the protocol required for the data collection. According to the study requirements (chapter 2), GPS technology is selected to collect data. It offers two ways to collect data: waypoints or tracking, and different possible settings. Even if, concerning existing data, tracks data appeared to be more relevant, the two ways to collect data should be explored. The objective is to give guidelines to collect the most suitable data. This is performed by testing different possibilities on the field and comparing the results. First, this chapter refers to the data sample collection on the field, next to the data analyses, to conclude with the guidelines for the collection methodology.

4.1 Field test: sampling collection

By applying the fieldwork exposed in chapter1, 12 tracks (6 on the way to go and 6 on the way back) have been collected in order to be analysed. The idea of those tests is to compare the devices, the individual interaction of users, different methods, different collection settings and their influence on the accuracy of the collected data.

Unfortunately a technical problem occurred during the fieldwork. The settings defined in Arcpad for the tracking data collection with a PDA coupled with a Bluetooth device haven't been taken into account. The four tracks have been all collected with the same setting of approximately one point every second, according the signal reception, instead of according four different settings. To solve this problem, the different settings have been made by "hand", by removing exceeding points. The result can be considered as correct because it is performed by removing the larger quantity of points without modify the quality of the collected data proper. In addition it allows to have five times more data. Indeed, each track has been modified to correspond to five new settings. That means that each tracks collected with the PDA and the Bluetooth GPS provide 6 sample datasets (5 new settings + the original collected one).

Concerning the Garmin device, the tracking setting is about one point every two seconds and it is not adjustable (Mehaffey and al.,2001).

The table 4.1 summarizes the modifications performed on the data and shows the final set of data available for the analysis. The total of data available is made of 32 datasets: 4 are waypoints, 4 are tracks with non adjustable settings, and 24 are tracks with different adjusted settings. The original collected data are represented in green and the new created ones in yellow.

Devices	Effective Setting	Modification	Final dataset (new setting)
Garmin	User A	No	Waypoint1
	User B	No	Waypoint2
	User C	No	Waypoint3
	Turn	No	Waypoint4
	Track automatic 1	No	Track go1
	Track automatic 2	No	Track go2
	Track automatic 3	No	Track back1
	Track automatic 4	No	Track back2
PDA+bluetooth	Track 1	Yes	Track (every 1 second)
			Track (every 1 minute)
			Track (every 10 minutes)
			Track (every 3 meters)
			Track (every 20 meters)
			Track (every 50 meters)
	Track 2	Yes	Track (every 1 second)
			Track (every 1 minute)
			Track (every 10 minutes)
			Track (every 3 meters)
			Track (every 20 meters)
			Track (every 50 meters)
	Track 3	Yes	Track (every 1 second)
			Track (every 1 minute)
			Track (every 10 minutes)
			Track (every 3 meters)
			Track (every 20 meters)
			Track (every 50 meters)
	Track 4	Yes	Track (every 1 second)
			Track (every 1 minute)
			Track (every 10 minutes)
			Track (every 3 meters)
			Track (every 20 meters)
			Track (every 50 meters)

Table 4.1: New collected data sampling.

Whatever the methodology selected, a remark has to be made about the compromise needed between the point density collected and the accuracy expected. Indeed, one limitation to take into account with GPS is the memory capacity of the devices. This particular point is often quoted as an important drawback of GPS devices, especially during long trips (*Brawn, 2003; Mehafrey and al., 2007*). This is a fact to consider, even if some improvements are done about it. The question is about finding the good balance between having enough points to get data accurate enough, considering the memory limitation, and the threshold accuracy value set in the introduction part. Moreover, too high point density is not always handy and necessary. Indeed, because of the inaccuracy of the devices, the results could be fuzzier with a too high point density. Then, some tests are performed on the field to first evaluate the two ways of collected data: waypoints or track, and secondly to find the most relevant parameters setting.

4.2 Data collection results

4.2.1 GPS users protocol: waypoints or tracks?

This part deals with the technique itself. The feelings of the user as well as the main advantages and disadvantages of the three methods used (waypoints, waypoints at turns and tracking) are detailed. Afterward those statements are supported by the data analysis calculations.

- **GPS waypoints collection**

At first hand, this method seems to require a lot of participation and attention from the user. Indeed it needs more work and it could be boring and restrictive. Moreover it is sensitive to variability between each person. The accuracy can decrease if the user forgot for a while to take waypoints.

On the other hand, this collection method is more adjustable to the situation. As it is controlled by the user, this latter can adapt the collection to the surrounding conditions. For example: taking less waypoints in straight line, more in turns or when the path is less clear, also stopping to collect points when the user go out of the track to take a picture. It is also possible to delete a wrong point directly if the walker takes a wrong path and has to turn back. Comments can also be added directly to the corresponding waypoint.

- **GPS waypoints at the turns**

This technique consists in taking waypoints only when the orientation of the path changes. It presents the main advantage of requiring few waypoints so less storage space. But the question about the accuracy level resulting from it arises. Indeed, the fact that sometimes it is difficult to evaluate when a turn occurred, can affect the result. It happens when the change in direction is very few but at the end important on a long distance. Those considerations allow to doubt about the sufficiency and the reliability of this method for the network creation.

Another consequence of the few numbers of points to take is the fact that the user could tend to forget to take one more easily. And in this case the consequence could be really important.

However, regarding only the navigation concern, this collection method may be clearer for the user by indicating only when there is a change in direction: no point or indication means to go on the same path. This last point can be of interest to provide navigation indications. Concerning the construction of a network, the data produced with this method appears to be not accurate enough, as shown in figure 4.1.

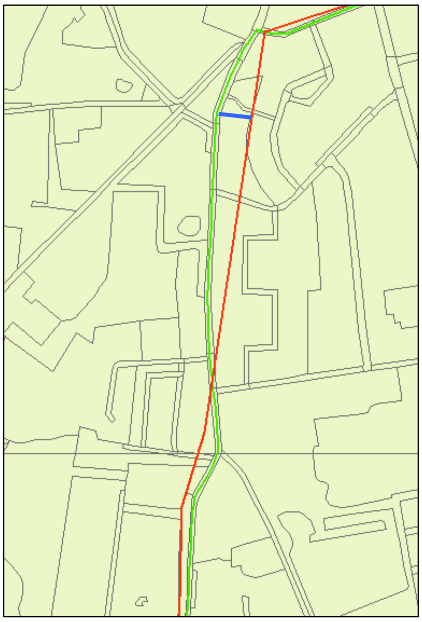


Figure 4.1: Display of the track collected with the turn method (in red) compared to the reference track (in green).The blue line shows a gap of 40 meters between the two lines.

• GPS tracking collection with a Garmin (non adjustable settings)

With this method a large quantity of points is collected (around 1 every two seconds, which correspond according to a walking speed of 4 km/hours to 1 point every 2,2 meters) and this kind of device has a really limited storage capacity. For the Garmin Etrex Summit, the tracking memory has a capacity of 10 000 points, then according to the collection frequency corresponds more or less to 6 hours of tracking, which is enough for short walking but can become insufficient for a trip of several days. Moreover when the memory is full the points collected at the beginning are erased automatically by the new ones. It happens also that sometimes the GPS signal interrupts and the user is not aware of it if he doesn't keep an eye on the device.

This device is a basic model and there is no possibility to choose setting or to have nice screen to display the result. But those drawbacks are only inherent to the model.

On the other side, this method is really easy going. For the user, there is almost nothing to do. It also ensures a quantity sufficient of points to give a reliable result.

• GPS tracking collection with a Bluetooth GPS+PDA (adjustable settings)

The comments about tracking data with Bluetooth GPS coupled with a PDA are relatively similar to the ones given for Garmin tracking. However, it presents more advantages. It is more comfortable and friendly to use because the tracking is displayed on time. Moreover it is also possible to display other maps or information underneath. In addition the settings can be chosen, so it allows a better adaptation to different situations and to the memory capacity.

➔ Then, based on the user point of view and the techniques characteristics, the automatic tracking may be the most comfortable and secure technique, even if it can generate some mistakes in the track.

Beside those advantages and disadvantages noticed by experimenters and inherent to the methods, some calculations are made to improve the assessment of the different data collection methods: waypoints or tracking, the waypoints at turns having let aside.

The results have been split in three tables. Table 4.2 offering the results for the data collected manually by waypoints with a Garmin device. Table 4.3 giving the result for the data collected automatically by tracking with a Garmin. And table 4.4 providing the results for the data collected automatically by tracking with the PDA+Bluetooth GPS. All the analysed data for the three cases concern the same path, to limit external influences and focus on the variation due to the methodology.

Point density (per 100m)	top10 deviation (m)	nb of polygons	mean area (m^2)	Smooth (from 1 to 5)	Polygon size
5,248	6,474	22	1850,14	4	3
6,834	4,683	25	1262,06	4	4
4,352	7,523	12	2529,38	3	3
5,478	6,227	19,67	1880,53	3,67	3,33

Table 4.2: Garmin waypoints evaluation.

Point density (per 100m)	top10 deviation (m)	nb of polygons	mean area (m^2)	Smooth (from 1 to 5)	Polygon size
9,079	6,448	27	1479,43	3	3
14,049	4,717	31	1028,92	3	4
9,741	7,812	13	2448,39	2	3
8,519	6,591	21	1514,35	2	3
10,347	6,392	23	1617,77	2,5	3,25

Table 4.3: Garmin tracks evaluation (with default settings).

Point density (per 100m)	top10 deviation (m)	nb of polygons	mean area (m^2)	Smooth (from 1 to 5)	Polygon size
85,194	7,679	27	1152,17	2	4
92,357	9,026	8	3072,44	3	3
76,312	6,793	147	226,673	1	3
90,694	7,663	13	1669,55	3	4
86,139	7,791	48,75	1530,21	2,25	3,5

Table 4.4: Bluetooth GPS+PDA tracks evaluation (with default settings).

The point density is higher during the tracking. And a fact directly linked to this observation is that the collected paths are less smooth when tracking. However the accuracy seems to be quite similar. The deviation polygons area appears to be a bit smaller when tracking even if the difference is not that large (1880,53 – 1530,21 = 350,32 m^2), and the number of those polygons is higher. This means that the inaccuracy is more located all along the track and swings along the reference line. So when tracking, the mistake are more spread in a random way.

Another remark concerns the variability of the accuracy. Indeed, the waypoint collection method present more variability than the tracking method (Tracking variance = 1.62 m / Waypoints variance = 2.06 m). This is due to the user influence inherent to waypoint collection method. In this case the different is not really important. It can be explain by the fact that the users were aware to be part of the experiment, even if they didn't receive any indications about how to collect waypoints.

In order to go deeper in the waypoint-track method comparison, the different results coming from the Garmin device have been compared to each other. Some cross-analyses are performed to see which percentage of deviation two tracks have in common. This allows to check if the deviation observed is more due to the method than to the device. To do this, the common polygons area between two tracks is computed. Then the value is used to calculate the percentage it represents for both tracks.

The analysis is run with the tracks to go on one hand and the tracks of the way back on the other hand, collected with Garmin devices in order to see the influence of the devices on the data collection:

	Dataset 1	Dataset 2	% of dataset 1 explained by dataset 2	% of dataset 2 explained by dataset 1
GO	UserA	Go1	94,77	96,57
	UserB	Go1	53,34	42,13
	UserA	Go2	40,32	51,45
	UserB	Go2	96,09	95,05
	Go1	Go2	41,92	52,5
	UserB	UserA	52,51	40,7
BACK	UserC	Back1	95,88	91,43
	UserC	Back2	52,11	49,73
	Turn	Back1	34,83	29,25
	Turn	Back2	57,77	48,56
	Back1	Back2	53,52	53,57
	UserC	Turn	31,02	35,22

Table 4.5: Garmin data cross-analysis.

UserA, userC, go1 and back1 used the same Garmin device, represented in blue in table 4.5. Whereas, userB, turn, go2 and back2 used the second Garmin device, represented in green. When two values coming from data collected with the same device are compared, the combination is underlined in grey in table 4.5. This highlights the fact that the highest deviation similitude (in red in table 4.5) are obtained by comparing results coming from the same device. In those cases, part of the deviation is explained by the device error and not by the method used. It is different for “turn-back2”, which have been collected with the same device but shows few corresponding. In this case, it points out the influence of the collection technique used.

According to the field tests and the analyses carried out here, there are not big differences between the data obtained from waypoints of tracking methodology; especially when the tracking is done automatically by the device without adjusting the settings. However, those conclusions about track and waypoint comparison can be different from a device to another one, and moreover from a selected setting to another. That’s why, it is essential to analyse and define the setting parameter to draw a relevant conclusion. The conclusion about the suitability of waypoints or tracks methods mostly depends on parameters setting and the internal characteristics of the device.

4.2.2 Waypoints collection setting

In order to find the most suitable setting for this methodology, some tests are carried out. Indeed the objective is to be the most accurate as possible considering the device memory capacity and the fact that some ways of collecting generate more “errors” in the data than others. For example stops during a walk can create cloud of points around a position when using a particular setting.

The waypoints settings are closely related to the user, to his way to collect data. Even if the deviation value remained under the accuracy limit, some variations can be noticed between the different users due to the individual influence (table 4.2). However all the collected tracks for this experiment remain suitable. It also has to do with the fact that the users were collecting the data seriously for an experimental purpose.

The conclusion would be that there are differences between users, however the results obtained can be suitable whoever the user. Then, it is correct to say that this method is suitable if the user keeps a certain point density that will be defined in the tracking collection settings part. However this method remains strongly under the user influence which makes it uncertain.

4.2.3 Tracking collection settings

In order to compare the data obtained with the tracking method, two analyses are carried out. The first one is a Pearson correlation analysis, also called "sample correlation coefficient", run with SPSS. It calculated the correlation number between each variable to measure the degree of association between them. A positive value for the correlation implies a positive association. A negative value for the correlation implies a negative or inverse association. In this case the variables are the factors calculated for the data analyses and presented in the chapter 2: the point density, the mean deviation, the number of polygons, the mean area of those polygons, and the smoothness of the track. The setting value is also integrated as a factor of the analysis.

The result provides useful information about which factors influence the others and what is the influence. To do this, the factors defined previously have been calculated for each datasets and used as inputs of the analysis.

Table 4.6 shows the result of the analysis performed on all the tracking data. For each correlation, two numbers are provided: the correlation value ("Pearson cor") between -1 and 1, which valuates the correlation, and the sigma value ("Sig.") between 0 and 1, which qualifies the relevance of the correlation value. A sigma value as low as possible means a relevant correlation.

		Point density	top10 deviation	Nb polygons	Mean area	Smooth	Setting
Point density	Pearson Cor	1					
	Sig. (2-tailed)						
Top10 deviation	Pearson Cor	-,256	1				
	Sig. (2-tailed)	,158					
Nb of polygons	Pearson Cor	,420(*)	-,176	1			
	Sig. (2-tailed)	,017	,334				
Mean area	Pearson Cor	-,248	,885(**)	-,278	1		
	Sig. (2-tailed)	,171	,000	,123			
Smooth evaluation	Pearson Cor	-,753(**)	,313	-,626(**)	,332	1	
	Sig. (2-tailed)	,000	,082	,000	,063		
Setting	Pearson Cor	-,544(**)	-,076	-,280	-,054	,314	1
	Sig. (2-tailed)	,001	,678	,121	,767	,080	

Table 4.6: Pearson correlation analysis table.

As expected, table 4.6 highlights a strong relation (red boxes) between the deviation value and the mean polygon area. It also points out an inverse correlation between point density and smoothness. In other words, when the point density increases the smoothness value decreases and collected tracks become fuzzier.

There is also an inverse correlation between the point density and the value of the setting. Logically, when the setting value increases (from one point every seconds to every 10 minutes or one point every 3 meters to 300 meters), the point density decreases.

Table 4.6, doesn't show the excepted correlation between the settings and the top10 deviation. It can be explained by the fact that the distance and time settings are analysed together. Then, to go in more detail in the correlation, the setting parameter is divided in two categories, one concerning the settings based on time and the other concerning the settings based on distance.

Table 4.7 presents the second steps of the Pearson analysis, which breaks down time or distance settings:

		Point density	Top10 deviation	Mean area	Smooth	Time	Distance
Time	Pearson Corr.	-,394(*)	,956(**)	,794(**)	,487	1	
	Sig. (2-tailed)	,025	,000	,000	,055		
Distance	Pearson Corr.	-,580(**)	,935(**)	,010	,672(**)	.(a)	1
	Sig. (2-tailed)	,000	,000	,970	,004	.	

Table 4.7: Pearson correlation analysis table, with settings factor detailed.

Table 4.7 shows the strong link between the settings, distance and time, and the mean deviation value. Moreover, it points out a relation between the distance setting and the smoothness of a track.

The purpose of those analyses is to define the most suitable setting to get the best data collection result. As settings directly influence the mean deviation, so the accuracy of the data, graphs 4.1 and 4.2 are drawn to visualise the linear regression between those two factors.

For this last point, it appears that more data was needed to get a relevant linear regression. So, they have been created by the same technique that the one used at the beginning of the chapter. In total 40 tracks have been generated from the original datasets, corresponding to10 new settings (4 x 10 = 40):

- 20 based on time setting: 1 point every 90 seconds, 120 seconds, 150 seconds, 180 seconds and 210 seconds.

- 20 based on distance setting: 1 point every 100 meters, 150 meters, 200 meters, 250 meters, 300 meters.

The detail of this operation is shown in appendix 2.

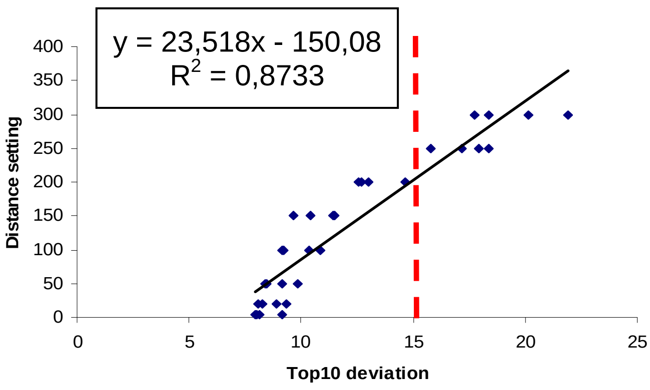


Figure 4.2: Top10 deviation according to distance setting.

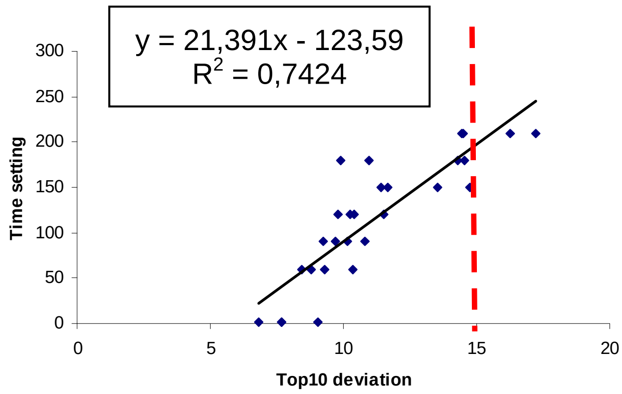


Figure 4.3: Top10 deviation according to time setting.

As expected, the deviation value increases with the increase of the time or distance settings. The threshold accuracy value of 15 meters fixed in the introduction is represented on figures 4.2 and 4.3 by a red dotted line. According to this limit, it is possible to visualise the settings corresponding the required accuracy.

4.2.4 Collection method recommendations

The experiments and analyses show that the data collection can be done by waypoint collection as well as by tracking. However, the tracking method seems to be the easiest to perform, especially if the trip is long. Moreover, the waypoints collection method asks for more attention and then can generate a result more unpredictable due to the individual influences. Then, it is less secure to collect this way.

Considering tracking, there are two cases: the device offers or not the possibility to choose the collection setting. In this experiment, only one device with a fixed tracking setting has been tested, but it can be considered as representative enough of the current devices available on the market. The results obtained with it were satisfying, even if the data were not really smooth. But the memory limited to 6 hours of tracking remains the main drawback concerning those devices, because they collect points with a quite high density and that is not possible to change it.

For the GPS with adjustable settings, the most suitable choice is based on the distance based setting because it doesn't create a cloud of points or fuzzy lines when the walker stops for a particular reason, at the opposite of time based collection. Figure 4.4 shows the different results for the same track collected with correspondent time and distance based settings (the correspondence is based on the walk speed of 4 km/hours). Indeed, a correlation between the distance setting and line smoothness has been previously highlighted.

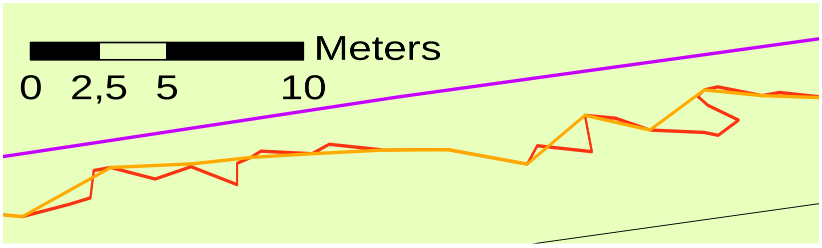


Figure 4.4: Difference between collection setting based on time (red) and distance (orange) compared to the reference route digitalised from top10 (purple).

According to the linear relation between the deviation value and the distance setting, the reasonable setting to use can be define around 1 point every 150 meters. This value shouldn't overstep 200 meters. This result can be expressed regarding the walking speed, which was around 4,2 kilometres per hour (1,17 m/ sec) during the experiment. That corresponds approximately to one point every 2 minutes. The maximum walking speed is 7,2 km/h. In this case, one point every 150 meters corresponds to 1 point every 1 minute 15 seconds. Then, if the setting is based on time, it has to take into consideration the walking speed of the user.

4.3 Conclusion

As proved by the field tests and the data analyses, there is no one unique fixed method for the GPS data collection. However the best way to collect data with GPS devices is defined as follows:

- Too many points are not necessary and can even be bothering by generating fuzzy lines.

- There is not a big difference between waypoints and tracking, if the user collects waypoint seriously. The main distinction resides in the user implication requirement. In this case, even if the testers didn't receive any indication about how to collect waypoints, they did it carefully because they were aware of the experimental issue. In a real situation the tracks collected by waypoints should have perhaps been less accurate. The result of waypoints collection is more unsecure, because based mainly on the user work, which is almost impossible to check. Then, the preferred method will be the tracking mode. This observation follows the one of the previous chapter advising to use preferably tracks data more than route data created out of waypoints.

The two previous statements can become contradictory. Indeed, when the tracking collection settings parameters can't be adjusted, the large quantity of point collected can create some memory capacity problems. But, even in this case, the tracking method remains the best collection choice according to the reliability of the result.

As the tracking is promoted, the corresponding setting is established considering accuracy and memory aspects. There is no one perfect setting either. This is mainly because the results are also different according to the device used. The final setting recommendation is based on distance with a recording frequency of 1 point every 150 meters. A frequency limit is set at 1 point every 200 meters. According to the devices, a more adapted trade off between data accuracy and memory capacity can be found.

V. Network development

The relevant web offered data have been selected in chapter 3 and new data have been collected according to requirements of chapter 4. This chapter 5 develops the second and third part of the methodology, namely the hiking network construction and segmentation. Before dealing with this, an extension of the data scope possibilities is offered. Next, the model to create the final dataset which will be used to build the network is developed. Afterwards, relevant network attributes are defined and the segmentation is performed. Finally, traversability rules are set and the network is build. This chapter ends with the validation of the created network.

5.1 Data overview

This part starts with a brief reminder of the previously selected data that will be used later on during the data processing. It also goes beyond the open mapping limitations by presenting another data source: geo-data features.

5.1.1 Data summary

In chapter 3, the GPS track data have been established as the most suitable data among the web offered data, according the availability, accessibility and compatibility criteria.

In chapter 4, the field tests and data analyses leaded to the conclusion that the most suitable collection method with a GPS is tracking based on a distance setting of 1 point every 15 meters. Then data have been collected this way.

For the network processing discussed in this chapter, those two kinds of data, the existing and the collected ones, are used. They have been selected and collected regarding the open mapping approach and the corresponding restrictions.

Even if those data are the most suitable data according to the open mapping approach, another potential data source is considered here in order to keep as broad as the choice of potential data. During the availability assessment, the geo-data appeared also to be suitable, but not of interest in this study because of their costs and use restrictions. However, they represent accurate and reliable data. Then it could be relevant to check their possible integration to the network and to keep them as a potential data if they appear to be suitable for the network construction.

5.1.2 Top10 vector data

Two kinds of data are offered by the top10 vector dataset: the road network and some footpath features. Indeed, road network features included in Top10 data can be considered as part of the hiking network since they can be used by walkers. Thus, all the components of the road network, except big arterial streets and highways, can be considered also as a part of the hiking network. The question is about how to deal with this road network. Two possibilities are conceivable: the road data can be added as another layer or integrate in the network itself.

One important point to consider is that some existing GPS data includes, as a part of the track, some road features. So using those data already includes some part of the road network in the walking network. However, there is a difference between the road presenting in the top10 data and the one collected by the user. This difference is explained by the device inaccuracy but also by the location of the user on this road when he was collecting the data. For example, the walker doesn't walk in the middle but at the side of the road. Then, it leads to a conflict between GPS tracks information which have been taken along the road and the road itself. It is true that it represents two different elements: the sidewalk and the road, but it still generates two or three lines for the same path. So, how to deal with this excess of information? According to this, it is better to keep the road information from top10 data as another layer of information, not adequate for the network construction. The road features will be progressively added to the network by the fact that there are included in GPS data, like normal walking paths.

Some footpath information is offered in the top10 data. At a closer look at the data features, this walking path information is mainly lines in the middle of nowhere (Figure 5.1). However, before using those data, their match for the project hiking path definition (chapter 1) must be checked. According to top10 data, public footpaths are defined by the following criteria:
- It is a path, paved or not, with a width inferior to 2 meters
- It is not a biking road, those latter are indicated by a traffic sign
- It is accessible by walkers but this criterion is not defined at all. Horse riding path are also classified as well as a footpath.


Figure 5.1 : Footpath features from top10.

Those criteria match the definition of a hiking path given in the introduction. Then, the interesting features can be extracted from top10 data and their integration in the hiking network can be tested.

Then, web offered data, collected data and footpath top10 features will be used to construct the hiking network. Those data have to undergo a specific processing. The firsts steps will consist in homogenize, correct and adapt them. Then, they will be compiled into a network.

5.2 Data processing

In this part, the data presented in the previous overview are used as a starting point to create the network. They are gathered in a final dataset that will be the base to build the network. The process shown in this part ends after the creation of this dataset integrating all the data perfectly together. The construction of the network out of it is done later on, after the definition and application of the segmentation.

5.2.1 Conceptual processing model

The idea is to aggregate the data in a coherent dataset. In other words: data from different sources ➔ dataset suitable to build the network for walkers using GPS. This is performed through different steps. As the data present different attributes, sometime almost no ones, and for sake of homogeneity, all the data attributes are deleted. Those latter will be defined further in the process. Then, the process starts with raw data, without any attribute.

To establish the conceptual model, the characteristics of the data have to be taken in consideration the first point is that there is very few information about those data especially about the collection method and the coordinate system used. This can generate some shift and error problems when data from different sources are compiled. All the data are not available in the same format either. So, the first concern is to homogenize those data.

To do this, the format and the coordinate system of the data have to be unified (1).

The second step concerns the improvement of the data. Indeed, the data may need some corrections (2). Tracks created out of GPS measurements are not always smooth or can present some discontinuities. Those defaults have to be removed or reduced as most as possible.

After those two steps, the data can be considered as homogeneous (a) and suitable (b) to be inserted in a network.

The next step of the network creation process aims to make the different data “fit” to each others. Indeed, they are sometimes several tracks collected for the same path and because of the accuracy of the GPS they are not perfectly at the same position so it gives a messy result. Moreover it is not useful to have several almost similar features for the same information. So the existing tracks have to be adjusted and the redundant information deleted (3). In this case a decision has to be taken to know which line will be kept or how to draw a new one out of the others. The last step is the aggregation of those different tracks in one final dataset (4), which will be used to build the network. All the process is summarized in the figure 5.2.

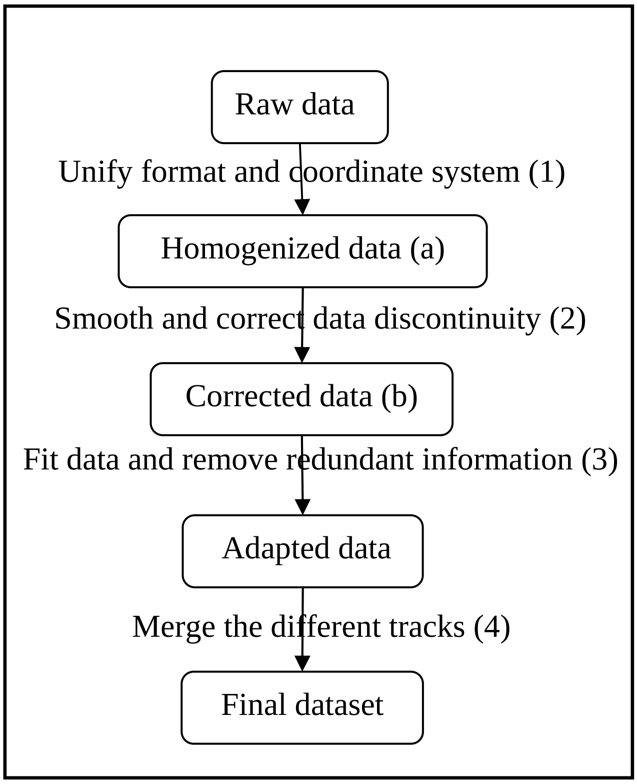


Figure 5.2: Conceptual model flowchart.

5.2.2 Logical model

The conceptual model described above is implemented in ArcGIS. Now, it has to do with the precise definition of each step define in the conceptual model.

As said, the first step consists in the data homogenization. Then, the data are exported in the same format and reprojected in the same coordinate system (1). According to the use of ArcGIS the format chosen for the data is the shapefile format. Concerning the coordinate system, the official Dutch projected coordinate system, Rijksdriehoekstelsel (RD new), is applied.

The second step required is the correction of the data in order to remove or decrease their internal mistakes (2), which can be due to outside points for example. This correction is firstly bring by the use of the integrate tool and after could be improved by another tool. The main objective of the integrate tool is to make the several data fit to each others (3), with, a tolerance set of 10 meters. That means that if two lines are spaced by less than 10 meters, they are considered as a same line that will be the mean of those two lines. With this tool there are no priority rules because all the tracks are considered as equals. This value of 10 meters has been chosen according the current GPS devices accuracy, the accuracy needed for this network and the result of some tries to find the most suitable value. As quoted before, the integrate tool, more than only making the tracks fit to each other, also bring a correction to the data. It makes the tracks smoother and moreover corrects some discontinuities. If this correction appears to be not sufficient enough to correct all the mistakes, a second correction tool (simplify or smooth tool) can be used later on to improve the result of the first one. After, the data need to be integrated again to fit perfectly.

Then the data are compiled to create the final dataset (4). For that, the merge tool is used. All the tracks are now in the same files but there is redundant information where the tracks are overlapping. Even if it is not possible to see it, this repetition of track is useless and can also disturb the good working of the network later on. It is then important to delete this unwanted information (5). For that, a new field is added and a common attribute value is given to each track. The dissolve tool is applied on this value. The cleaning tool can also be used. Finally, a topology based on a no overlapping rule can be created to check that there is no more redundant data (6).

After having performed all those steps, a final dataset is available, which consists in one complicated polyline without segmentation or attribute.

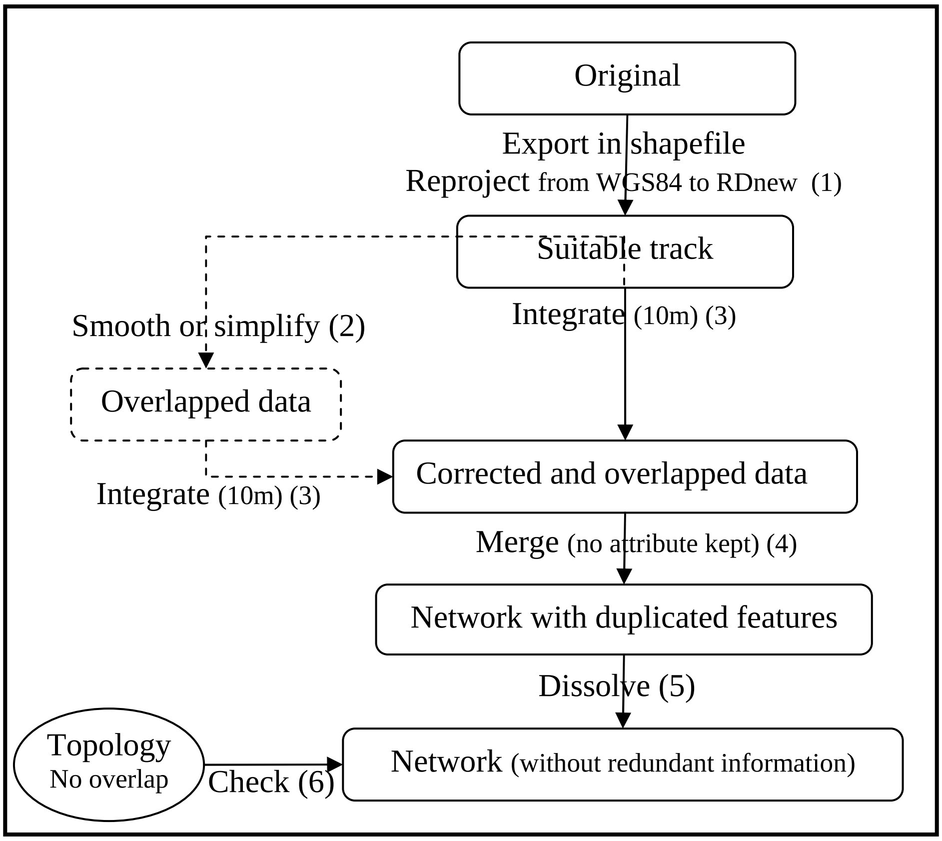


Figure 5.3: Logical model flowchart.

5.2.3 Implementation model: Application in the Veluwe area

The first step, after having deleted the attributes of the original data, is to export the data in the same shapefile format in ArcGiS. Then the reprojection is performed. Each data file is reprojected from WGS84 or Amersfoort datum, according to the original coordinate system, to RD new. For the conversion from WSG84 to RD new, the corresponding geographic transformation called "from WGS84 to Amersfoort" is applied. The one used in this study is the oldest transformation established; currently there is a latest one. It results in projected files presenting a georeference and format homogeneity. The data can now be integrated, corrected and merged.

Before to be integrated, the projected data are copied in order to save them. Indeed, the integrate tool application modifies directly the data erasing the previous original information. The data copies are modified by the application of the integrate tool to fit to each other with a tolerance of 10 meters. This allows to correct the tracks. With the dataset used in this study, the correction bring by the integration is sufficient. Then, all the modified features are merged in one file: the final dataset grouping all the different tracks together.

The last step deals with the cleaning of this final dataset. The point is to remove the redundant information and to remove all the possible segmentation remaining from the original tracks data. A copy is made first, for the same safety reason than in the previous stage. A new field (name: "diss") is added to the copied data and then calculated to attribute the same arbitrary value (expression: "diss"=1) to each segments. Based on this value a dissolve operation is applied to obtain final dataset made of one unique polyline.

This implementation has been carried out by constructing models with the model builder application of ArcGIS (Appendix 3).

As explained, the integrate tool makes the data fit to each other by modifying them. Then, it is important to check the impact of this operation on the accuracy of the data. The mean deviation technique used to evaluate the data can also be used to observe the influence of the integrate tool application on the track, to check if it doesn't generate a higher inaccuracy.

Length (m)	deviation area (m^2)	mean deviation (m)	deviation area (integrate track) (m^2)	mean deviation (integrate track) (m)	Difference Original/ integrate
10486,105	31978,015	3,04956	30681,585	2,92593	0,12363
30721,546	140549,846	4,57496	131067,156	4,26629	0,30867
42565,939	152550,092	3,58385	149890,051	3,52136	0,06249
20313,894	96357,183	4,74341	97170,671	4,78346	-0,04005

Table 5.1: Integration effect on the mean deviation.

According to the results of table 5.1, the integrate tool doesn't affect significantly the existing deviation of a track. The shape of the existing tracks is improved by the correction brought by tool but the navigation information quality is not decreased, even improved a bit.

5.3 Network segmentation

The idea is to divide the network in useful part, suitable for the application of the network later on. Many questions arise with the decision to take about how to segment this network. The first question concerns the choice of the segmentation to use, and what could be the smallest division relevant for this network. Indeed a path is not like a road, especially in non urban area, there are not well defined beginnings and endings. This segmentation question can start to be solved by answering another question about which suitable attributes can be used to describe this network?

5.3.1 Descriptive attributes

No original data attributes have been kept because none of them were similar and almost not attributes were useful and relevant.

Figure 5.4: Example of attributes describing the GPS web offered data.

The attributes are required to identify a track and to give details about it. Then, based on some examples found on the Internet (Figure 5.4) or according to the standard ones use for official geo-data, the main relevant and suitable attributes are selected to describe the data for the hiking network application purpose:

• **Identification**: an individual identification number, added to each segment. The identification attribute, is a number automatically assign to a feature element by ArcGIS.

• **Name**: an unique identification name. The identification number could be enough to identify the segments but, as in the road network, giving a name is more clear and friendly. Then, each individual network elements should have one text name. It can be a combination of attributes like: “category-id-municipality”

• **Category**: the path category depends on the land use the path belongs to. The three possible categories are: built-up, agricultural or nature area. This information comes from the Dutch national land use dataset, lgn5 (Figure 5.6: category level)

• **Update**: it is the date of creation or of updating of the segment.

The update is an essential indication to bring information about the possible path changes that could occurred. In the definition of what can be considered as part of the hiking network, it was specified that only maintained paths are of interest. But the sensitivity to changes of this network is one of the particularities to deal with. Then this information can be used to know if a segment has been digitalised long time ago and if some changes could appear. This information has to be linked to the category. Indeed, a path is more likely to change in natural area than in built-up area. This information allows to know which track should be checked, because a network is never fixed.

If this date is not available for the existing data, it will be replaced by the date of addition in the network like what is done for the integration of the collected data.

• **Comments**: if a particular information has to be added about the path.

The comments attribute is a miscellaneous column to add some interesting information, like the presence of animals, that are of interest but not possible to specify somewhere else. The seasonality aspect quoted in the introduction can be integrated in this part as warming information.

• **Municipality**: the belonging municipality of the segment.

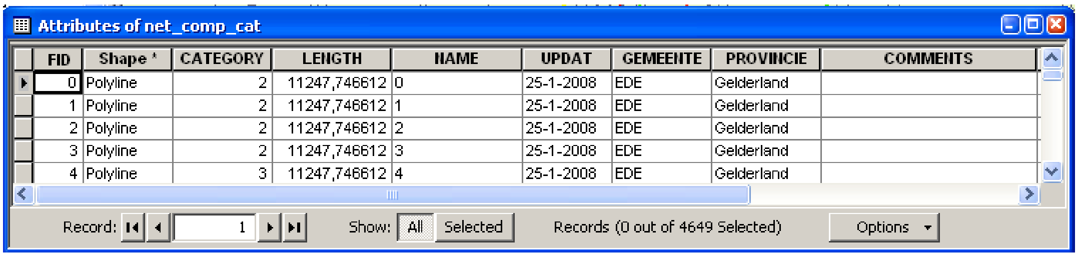
• **Province**: the belonging province of the segment.

Municipalities and provinces indicate the location of the segment. It brings a first segmentation to the network according spatial location. To find those attributes, the national Dutch territory division dataset is used. It refers to the organisation of the country, for example in the French context it would have been “ région”, “département” and “commune” divisions.

• **Length**: the length of the segment

The length is a basic information in a network for many calculation and path computation. This value is automatically calculated when the dataset is integrated to the geodatabase.

All those attributes are added to the final datasets (Figure 5.5), obtained from the previous data processing, dividing it in several segments.



Attributes of net_comp_cat

FID	Shape *	CATEGORY	LENGTH	NAME	UPDAT	GEMEENTE	PROVINCIE	COMMENTS
0	Polyline	2	11247,746612	0	25-1-2008	EDE	Gelderland	
1	Polyline	2	11247,746612	1	25-1-2008	EDE	Gelderland	
2	Polyline	2	11247,746612	2	25-1-2008	EDE	Gelderland	
3	Polyline	2	11247,746612	3	25-1-2008	EDE	Gelderland	
4	Polyline	3	11247,746612	4	25-1-2008	EDE	Gelderland	

Record: 1 Show: All Selected Records (0 out of 4649 Selected) Options

Figure 5.5: Network attribute table.

5.3.2 Network building

The segmentation of this network is made of three levels (Figure 5.6). The first division is done according the location. The network is segmented according to the province and municipality divisions. However, this division is not sufficient, the scale is too large. Then, a more detailed sub-segmentation is required. This latter is built on the category attribute coming from the land use information. Even if this sub-segmentation is relevant for this network and will be used in computation, it is not enough either. To solve this, a preliminary segmentation can be performed beforehand. This first segmentation is based on intersections. Each intersection is considered as a node between arcs, namely the end and the beginning of a feature element. This is performed with ArcInfo.

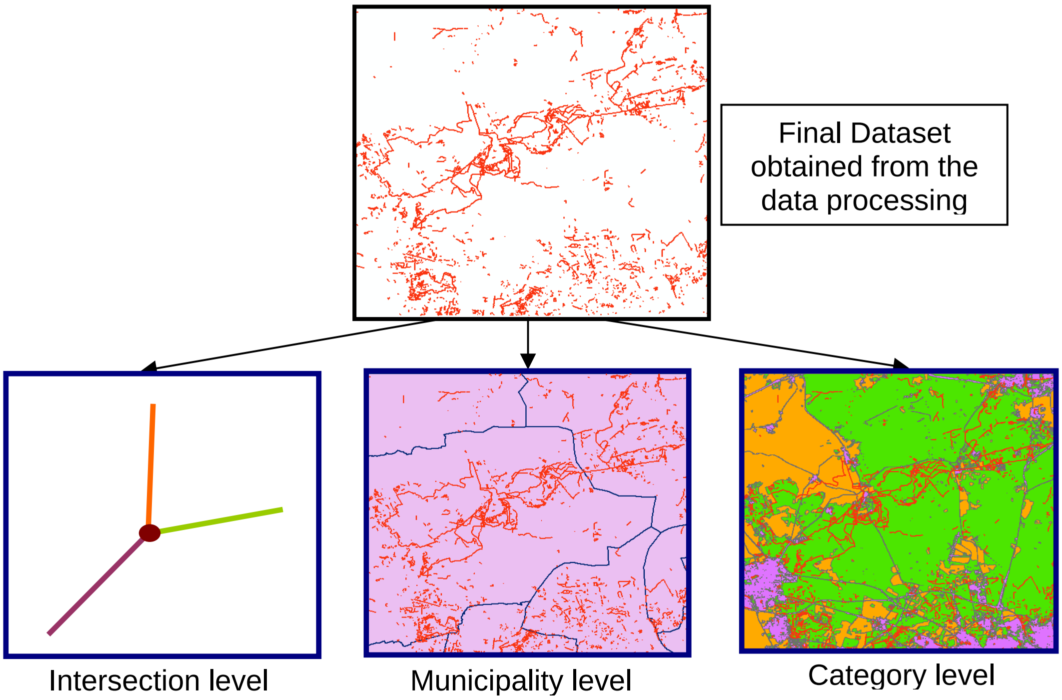


Figure 5.6: Segmentation level.

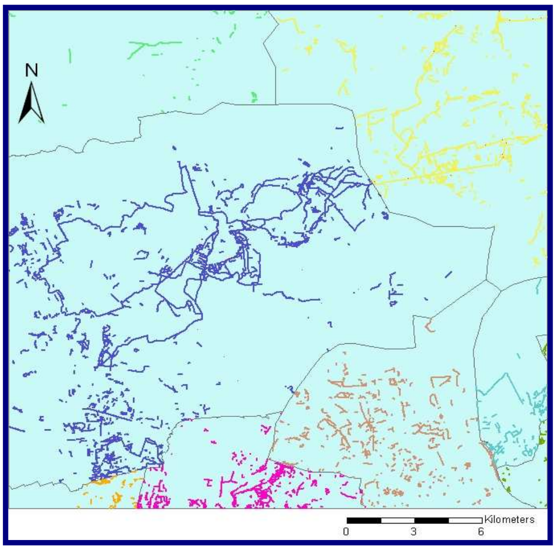


Figure 5.7: Network displayed according to the municipality segmentation.

Now that the segments are defined, it is interesting to wonder about the quality and usability of sub-segments. Concerning the second and third segmentations, they are based on existing datasets: territory administrative division and land use. So, the accuracy of those segmentations is ensured by the accuracy of those datasets.

As this project deals with pedestrian navigation, the construction of the network is simpler for some aspects than the road network. A network is defined by attributes, which are the properties elements that control traversability over the network. One property of the network to define is the restriction regarding this traversability aspect. In this case, because of the walker freedom, there is no use to specify a restriction according to the traffic direction. All the paths can be travelled in either one direction or the opposite one. It is also not necessary to assign particular turn feature restrictions for the same reasons (Figure 5.8). Thus, the network is simply designed with no particular restrictions. However some attributes can be used as hierarchy criteria during a path computation, the category for example can be used by walkers who want to avoid city.

For a hiker, the main restriction is based on the difficulty of the walk. Then, a difficulty level attribute could be added to improve the relevance of the network.

Figure 5.8: Scheme of the hiking network building rules.

Based on those rules, the hiking network is built.

5.4 Network validation

5.4.1 Path computation

In order to check the final result, the network is built in ArcGIS and a test route is computed (Figure 5.9). Four walkers navigated through the area followed this route. During the trip, they experienced the network and then gave their comments about the accuracy and the suitability of the network for hiking purpose. The test group was made of four different walker types. There was one person who already knew the area and was used to read maps, two persons who didn't know the area but were used to maps also, and one person not familiar with the area and maps either.

This route has been chosen because it contains the three different possible land use categories. It makes use of information coming from different data sources and passes by intersections that offers to the user several possibilities and generated navigation misguidance.

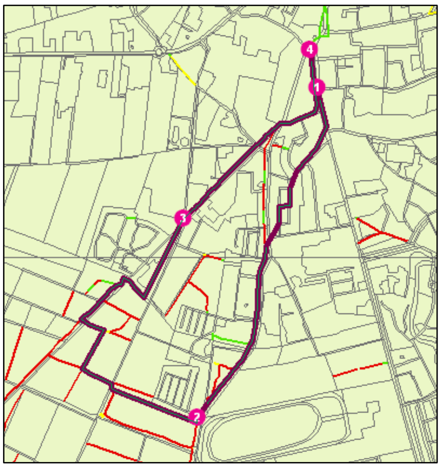


Figure 5.9: Automatic computed route.

5.4.2 Field validation

According to the network assessment method described in chapter 2, a navigation test was performed to check the sufficiency of the given network for hiking purpose. The four testers, experimented or not with the navigation, were guiding by successively. The result consists in the achievement to follow the computed route and in the comments received from the testers.

The main point was the success in travelling correctly the computed route. According to the testers, the network appeared to be sufficient to navigate through the area and no negative remarks have been expressed. However, some comments have been stated about the difficulty to navigate, more important in some particular areas. It concerns high path density area, like residential area where there are many small paths between houses, and forest area where the paths are not always clear or maintained. They can disappear and new ones can be created. In this case it was really helpful to have an overview of the all network. It allowed us to see other possible routes and helped us in the way finding, which would not have been possible with only travelled route information. Then a particular attention has to be paid for this kind of area concerning the accuracy and the updating of the data.

5.5 Conclusion

The objective of this part was to gather the different data in a suitable hiking network. To do this, homogenization and integration operations have been carried out in order to make the data compatible. Afterwards they have been merged in a coherent network. Attributes have been chosen to describe the network and to create the segmentation. This latter is based on location, land use and intersections. The validation test drove to the conclusion that the created network was suitable for the navigation purpose, even if some area are more confusing than others.

Then, the network constructed out of GPS existing and collected data is a good start to extend and improve the hiking network. Moreover, the use of geo-data is also possible to support the construction and improvement of the hiking network.

VI. Conclusion, discussion and recommendation

Regarding the results obtained and according to the objectives formulated at the beginning, some conclusions are drawn about the methodology to improve the hiking network offered in this report. Then some critical points of the work are discussed, and a comparison with the French hiking network created by IGN (National Geographic Institute) is presented. Finally, according to the unexpected problems occurred and the ideas for further development raised, some recommendations are expressed

6.1 Conclusion

This thesis project aimed to improve hiking network for GPS users through the establishment of a suitable methodology. The open mapping approach has been adopted to achieve the objectives. This choice generated the potential use of existing data and the collection of new ones with GPS devices in order to keep the method cheap and technically accessible. Three main questions had to be addressed: is the existing data usable for the construction of a hiking network, how to collect new suitable data with the GPS to support network construction, and how to gather those data in a final hiking network?

➔ **Improvement of the hiking network for GPS users by…**

… Exploring available hiking data

Even if a lot of data are available, not all are usable. This study focused on digital data, classified in three data types: GPS data, digital maps and geodata. Those data are characterized by a broad heterogeneity, even inside of the same category. There is no standardization. The second essential observation was that there is really few information about the data itself. Thus it is really difficult to evaluate and use them. According some accessibility and availability criteria, the GPS data appeared to be the most suitable ones, especially the GPS track data. They have also been declared compatible.

➷ As proved by the availability, accessibility and compatibility assessment, the web offered track GPS data are usable for the improvement of the hiking network

… Collecting new data

It is about finding the most suitable way to collect data with GPS devices. To do this, two techniques, waypoint or track, and their corresponding setting needed to be defined through several tests analysis. It resulted that there is no one unique fixed method. The both collection techniques, waypoints and track, have proved to be relevant for the purpose. However, taking in consideration the user's influence, the tracking is a more secure way to collect because it depends less on user reliability.

Then this method is preferred to the waypoint collection. The setting advised is based on distance with a frequency of 1 point every 150 meters, with a threshold value of 1 point every 200 meters.

➷ As proved by field tests and data analysis, there is no one unique method but the most suitable way to collect new data with a GPS is tracking 1 points every 150 meters.

… Processing and segmenting the network

For this last point, a network has been built out of the suitable existing data and the collected one. To open the view of this project, top10 features has also been considered. The road network defined as information of interest for the walker can be used as another information layer and the footpath information can be integrated in the network process. The segmentation has been based on the location, municipalities and provinces boundaries, and on land use category, build-up, agricultural and natural area. These divisions have been preceded by the creation of new segments at each intersection.

➷ The GPS offered and collected data can be processed to improve the hiking network and this methodology can also integrate geo-data.

The objective was to improve hiking network for GPS users. Regarding to the result, this report offers a methodology to go from several sources of data to a suitable segmented network, validated in the field. Then it is correct to say that the objective has been achieved. In this case, it has been performed according to the open mapping approach, which was one of the possible ways to explore the issue. This work brings a sound starting point for further developments to carry on improving the hiking network for GPS user.

6.2 Discussion

6.2.1 Main discussion points

➔ The datasets used as reference are relevant but bring their uncertainty to the result

To evaluate the existing and the collected data, the mean deviation value has been calculated. However this value has to be used carefully. Indeed the reference is based on the top10 vector dataset, which presents an assumed error of 2 meters (*BUREN, J VAN et al. 2003, 2*). Moreover the reference route has been drawn in the middle of the travelled path or road, which can be also 2 or 3 meters difference from the real location of the walker. Also small paths between houses have been travelled and they are not always represented on the top10 data, so it generates the creation of an approximate reference at some places. All those points lead to an uncertainty brought to the mean deviation computation. It would have been more exact to record the travelled route with highly accurate measuring instruments like RTK, to produce the reference track. However, several kilometres have been travelled and using those instruments would have been time consuming and not really handy to carry. The calculation of mean deviation values compared with the use of RTK measurement as reference has already been performed in the Achterhoek (Bulsink and al., 2007). It pointed out the reliability of the deviation result even if there is still an uncertainty.

This remark highlights another point to consider. In the mean deviation calculation, there is an uncertainty coming from the reference used, as exposed previously, but there is also an uncertainty coming from the collection step. This can explain some variations observed in the accuracy of the data. Indeed, the different positions of the different walkers on the path during the walk compare to the chosen reference can influence the result. To be perfectly similar the tracks should have been collected with the same device, even if it is the same model, and by people walking exactly at the same position compare to the reference used. It has been noticed that the two users who walked with exactly the same GPS device generated a higher deviation, but there are not enough measurements to conclude about it. Then the user behaviour as well as the device inherent properties lead also to a particular uncertainty. However, the computed results remain relevant for the data assessment purpose.

➔ Collection settings should be reconsidered according the density of the surrounding

The setting indication of 1 point every 150 meters was formulated for the data collection methodology. This information has to be considered in regard to the complexity of the area. In fact, if the surrounding area is confusing because of many tracks or a lot of turns, the number of collected point should be increased to adapt and be more accurate. Indeed, it has been quoted several times in this report that some complicated areas, like residential or forest areas, appeared to be difficult to navigate for hikers. Then a relation between complexity of the surrounding and the setting indication value need to be established. Some tests can be carried out to break down and represent this relation. In this particular case, the technique of taking waypoints at turn could be reconsidered.

➔ Freedom of a walker to cross a square area fencing notion

The last remark concerns a particularity due to the fact that this work is dealing with pedestrian network. The freedom of the walkers generates many different possibilities how a square can be crossed, barriers like greens and fields can be passed without following a certain path (*Gartner and al., 2005*). Thus it can't be considered as a normal path. However, when users provide their tracks, there is no way to know if a part of a route corresponds to an open square area, so they are integrated as a normal segment of the network. They should be considered in regards of the fencing notion, which refers to a delineate area with a network arriving point and a departing point and many ways for the user of the network to connect them.

6.2.2 Comparison with French example

Description of IGN and Georando

The National Géographic Institute (IGN) is a governmental company which produces, maintains and distributes the geographic information of reference in France. In that capacity, it established the French cartography in the form of paper maps, digital maps or geodatabase. In this context, they are leaded to show some pedestrian paths on the maps or in the geodatabase, according to the scale. Even if they represent them, they didn't create them.

Throughout paid partnerships, the hiking routes are displayed on some maps, at the top of the other data (Figure 6.1). Those routes, which are the combination of walking paths and roads, are not the property of the IGN but they are owned by their authors. For example, the "Fédération francaise de randonnée" which owns the brands "grande randonnée" (GR) and "petite randonnée" (PR).

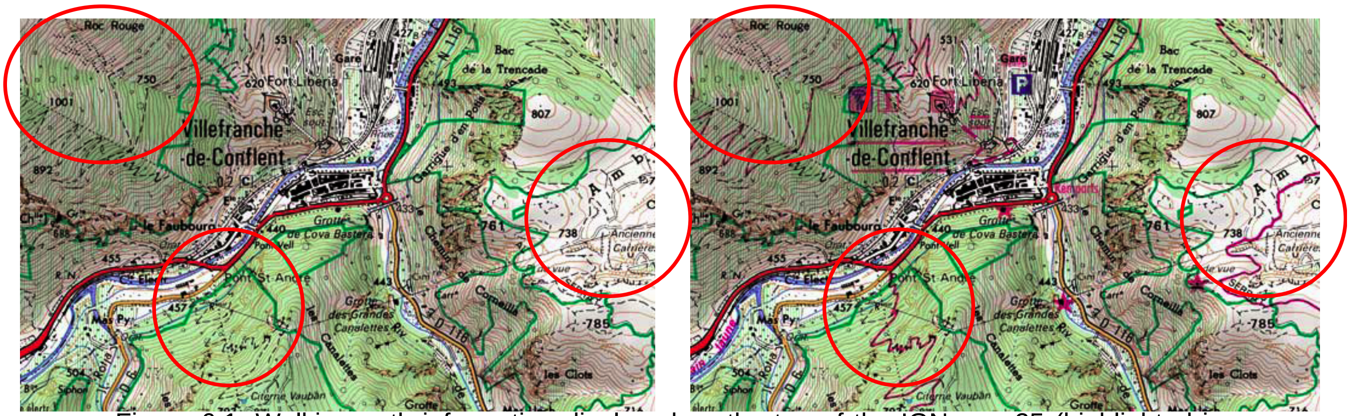


Figure 6.1: Walking path information displayed on the top of the IGN scan25 (highlighted in violet on the tourist version).

The Georando DVD is a commercial product co-edited by IGN and the company Star Apic. It combines geo-data network and software to plan hiking trips on a computer. Especially, it allows to drawn or to compute routes based on the 1:25 000 maps and aerial photos, as shown in appendix 4. It also offers the possibility to calculate the profile and the differences in altitude, to visualise it in 3D (Figure 6.2) and to export it as GPX files on a PND, PDA or Smartphone connected to a GPS devices. In France this kind of product is also marketed by Bayo and Memory Map editors.

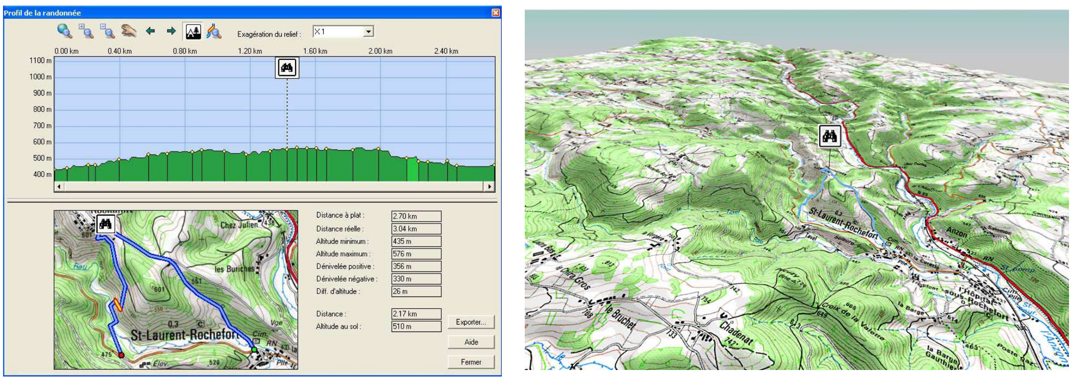


Figure 6.2: Examples of hiking route visualization in Georando.

Until now, the hiking routes were almost only available in a paper format, through paper maps or books. The organizations in charge of managing and maintaining the hiking paths just start to use the digital format. The hiking routes are digitalised based on 1:25 000 IGN maps, which have an accuracy of 2,5 meters.

Similarities and differences

The French example is a totally different way to deal with the same objectives. The approach chosen in this thesis work, was the open mapping slant. In Georando, it has been treated from a commercial point of view. The main difference is based on the accessibility. The software Georando is lucrative and the data are authorised, so the usability is under restraints. On the other side the methodology proposed in this project aims to create a network which is freely accessible and usable.

The common point is the use of existing data. Indeed both methodologies use data already created by others. However, for the commercial point of view, those data are bought against payment. More existing data can be used because the cost is not a limitation anymore and also because most of the data already exist. Then, there is no need to collect new one. With the open mapping perspectives, the costs have to be kept as low as possible then it is not possible to buy data. The solution is to collect what is not available or accessible. However the existing data used in both methodologies are not the same. For Georando, they come only from organizations that digitalise them by drawing on a based map whereas the ones used in this project consist in GPS data.

To conclude, the question about how to consider the road network is treated in the same way. The parts of the hiking routes which correspond to a road are added as a part of the hiking network, otherwise the road network can be displayed underneath in the 1:25 000 maps.

6.3 Recommendation

The main leads given in the discussion part were dealing with the data uncertainty and the analysis of density surrounding influence on the advised collection settings. Concerning this latter, it should be recalculated according to the remarks and tested in a more relevant area. Beside those ideas coming from methodology limitations noticed during the research work, some ideas to go on with the presented work can be formulated.

Further development and maturing of the open mapping application

The combination of the web offered GPS data and the collection of new ones proves to be suitable for the improvement of the hiking network. The presented methodology offers a good starting point for further development and maturing of the open mapping application. The network process methodology needs to be extended to allow different users to participate in the network improvement by adding their collected data. This can be done through a Internet application based of the wiki principle, like for the open street map project ([url18]).

Looking for GPS accuracy improvement by using hybrid method or applying Egnoss correction

The GPS technology has been chosen in this project because it corresponded to the cost and the accessibility sought. One drawback inherent to this technique has been quoted: the decrease of the accuracy, especially under vegetation cover. It could be of interest to go into detail with this topic. Indeed, this loss of accuracy could be balanced by the use of inertial technique. But this option would restrict the extension of the methodology to open mapping. The development of Egnos corrections (*Naus, 2007*) directly integrated in the new devices, which will become more and more available in the future. EGNOS (European Geostationary Navigation Overlay Service) is a satellite based augmentation system. It is intented to correct GPS, GLONASS and later Galileo systems by improving at the same time the reliability and the accuracy of the signals. It consists of three geostationary satellites and a network of 40 ground stations, spread all over Europe. According some tests the GPS accuracy is about to two meters with Egnos corrections. Then, it could be a way to explore in order to improve the general GPS data accuracy.

Development of a formal data model for the data standardization

Another point mentioned during this work was the lake of data standardization. Some international projects like the European “Walk on Web” ([url9]) and the American “Trail Data Content and Data Transfer Standard” ([url8]), aim to facilitate the sharing of hiking routes and walk paths data. It could be usable to concentrate more particularly on the study and establishment of a suitable formal data model.

Difficulty level integration

As quoted during the establishment of the rules for the network construction, the main restriction for a hiker is the difficulty of the walk. Then, it would be relevant to create a model in order to add this essential information to the network. Some leads are offered in appendix 5. However, this aspect should be developed.

References

• ANDRIENKO N., ANDRIENKO G., GATALSKY P., 2003. *Exploratory spatio-temporal visualization: an analytical review.* Journal of Visual Languages and Computing 14, 2003, Pages 503 - 541.

• ARLINGTON MASTER TRANSPORTATION PLAN., 2006, Pedestrian Element. Virginia, USA, November 2006.

• ASAKURA Y., IRYO T., 2007, *Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument*. Transportation Research Part A: Policy and Practice, Volume 41, Issue 7, August 2007, Pages 684-690.

• ASLAN I., KRUGER A., 2004. *The Bum Bag Navigator (BBN): An Advanced Pedestrian Navigation System*. Workshop on Artificial Intelligence in Mobile Systems (AIMS), UbiComp, 2004.

• BAUS J., KRUGER A., WAHLSTER W., 2002. *A ressource-adaptive mobile navigation system*. Conference on Intelligent User Interfaces IUI, San Francisco, USA, 2002.

• BREGT A., BULENS J.D., GROTHE M.J.M, JANSSEN P.A.L.M., OOSTEROM P.J.M.van, QUAK W., REUVERS M., RINK M.A. de, SMITS P.C, 2006. *Framework van standaarden voor de Nederlandse GII*, versie 1.1.

• BULSINK D.J., 2007. *Validating GPS-track descriptions for walking routes.* Wageningen University and Research Centre - Alterra, Centre for Geo-Information, The Netherlands.

• DOYLE D., 2005. *Accuracy test of Consumer Grade GPS Receivers*.
[url1] http://www.doylesdartden.com/gis/gpstest.htm, 31 July 2007.

• D'ROZA T., BILCHEV G., 2003. *An overview of location-based services*. British Telecom technology journal, vol. 21, no 1, pp. 20-27.

• GABAGLIO V., 2003. *GPS/INS integration for Pedestrian Navigation.* Geodätisch-geophysikalische Arbeiten in der Schweiz, Zulrich, Switzerland, Vol. 64, 161 pages.

• GARTNER G., RADOCZKY V., RETSCHER G., 2005. *Location technologies for pedestrian navigation*. Austria.
[url2] http://www.gisdevelopment.net/magazine/years/2005/apr/location.htm, 17 April 2007.

• GARTNER G., 2004. *Location-based mobile pedestrian navigation services: the role of multimedia cartography.* ICA UPIMap, Tokyo, Japan, 2004.

• GNATEK T., 2004. *For Wandering Tourists, Help From on High*. The New York Times, 10 June 2004.
[url3] http://www.hci.cornell.edu/news/nyt1.pdf, 19 April 2007.

• GIJZEN S.W.J.G.M., 2007. *Validating track descriptions for recreational purposes: comparing trimble RTK and Bluetooth GPS measurements.* Wageningen University and Research Centre - Alterra, Centre for Geo-Information, The Netherlands.

• GILLIERON P.Y, LADETTO Q., 2002. *De l'évolution du GPS à la navigation pédestre*. Flash Informatique, informatique mobile, spécial été 2002.
• HAZEU G.W., 2006. *Land use mapping and monitoring in the Netherlands (LGN5).* Wageningen University and Research Centre - Alterra, Centre for Geo-Information, The Netherlands.

• LAHM J., 2007. *GPS Education Resource.* GPS.
[url4] http://www.gpseducationresource.com/gps_overview.htm, 31 July 2007.

• LAMMEREN R. van, 2006. *Digitale Wichelroede*. Wageningen University, 17 pp., RGI-156.

• MEHAFFREY J., YEAZEL J., 2001. *Garmin's eTrex Summit-Upgraded eTrex has some Interesting New Features*
[url5] http://www.gpsinformation.net/main/etrexsum.htm, accessed 12 December 2007.

• NAUS P., 2007. *Location Based Edutainmen, using the physical movement of a GPS receiver to trigger movement in a visualization of underground object.* Wageningen University and Research Centre - Alterra, Centre for Geo-Information, The Netherlands.

• O'CONNOR A., ZERGER A., ITAMI B., 2005. *Geo-temporal tracking and analysis of tourist movement*. Mathematics and Computers in Simulation, Volume 69, Issues 1-2, 20 June 2005, Pages 135-150.

• OREG S., NOV O., 2007. *Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values*. Computers in Human Behavior, 2007.

• PHILLIPS M.L., HALL T.A., ESMEN N.A., LYNCH R., JOHNSON D.L., 2001. *Use of global positioning system technology to track subject's location during environmental exposure sampling.* Journal of exposure analysis and environmental epidemiology, 2001, Pages 207-215.

• QUIROGA C.A., BULLOCK D., 1998. *Travel time studies with global positioning and geographic information systems: an integrated methodology*. Transportation Research Part C: Emerging Technologies, Volume 6, Issues 1-2, February 1998, Pages 101-127.

• ROOIJ B.J.R. van, 2007. *Validating track descriptions for recreational purpose:, from gps stream to waypoints.* Wageningen University and Research Centre - Alterra, Centre for Geo-Information, The Netherlands.

• ROUPIOZ L., 2007. *Validating digitalized routes for recreational purposes*. Geo information and Remote sensing integration, Wageningen University, 4pp.

• RUPPRECHT W.S., 2007. *Post SA GPS Accuracy Measurements.*
[url6] http://www.wsrcc.com/wolfgang/gps/accuracy.html, 31 July 2007.

• SHOVAL N., ISAACSON M., 2007. *Tracking tourists in the digital age*. Annals of Tourism Research, Volume 34, Issue 1, January 2007, Pages 141-159.

• STEINIGER S., NEUN M., EDWARDES A., 2006. *Foundations of Location Based Services Lesson 1*. CartouCHe1 - Lecture Notes on LBS, V. 1.0
[url7] http://www.geo.unizh.ch/publications/cartouche/lbs_lecturenotes_steinigeretal 2006.pdf, 17 April 2007.

• WEIMANN F., TOME P., WAEGLI A., AICHHORN K., YALAK O., HOFMANN-WELLENHOF B., 2007. *SARHA – Development of a Sensor-Augmented GPS/EGNOS/Galileo Receiver for Urban and Indoor Environments.* 7th Geomatic Week, Barcelona, Spain, 20-23 February 2007.

• WINTGES T., 2002. *Geo-Data Visualization on Personal Digital Assistants (PDA*). Maps and the Inter-net 2002, Volume 60, 2003, pp. 178-183.

• YOSHIMURA T., NOSE M., SAKAI T., 2006. *High-end GPS vs. low-end GPS: comparing positional accuracy in forest environment.* Graduate School of Informatics, Kyoto University, Japan

• ZEIMPEKIS V., GIAGLIS G.M., LEKAKOS G., 2003. *A Taxonomy of Indoor and Outdoor Positioning Techniques for Mobile Location Services.* SIGecom Exchange, Vol. 3, No. 4.,2003, Pages 19-27.

• [url8] http://www.fgdc.gov/standards/projects/FGDC-standards-projects/trail-data-standard/trail-data-standards, Federal Geographic data Committee, accessed 15 January 2008.

• [url9] http://www.walkonweb.org, accessed 5 June 2007.

• [url10] http://www.tracegps.com, accessed16 June 2007.

• [url11] http://www.gpsies.com, accessed 20 July 2007

• [url12] http://www.gpstrack.nl, accessed 20 July 2007

• [url13] http://www.gps-info.nl, accessed 21 July 2007

• [url14] http://www.ontrack.nl, accessed 22 July 2007

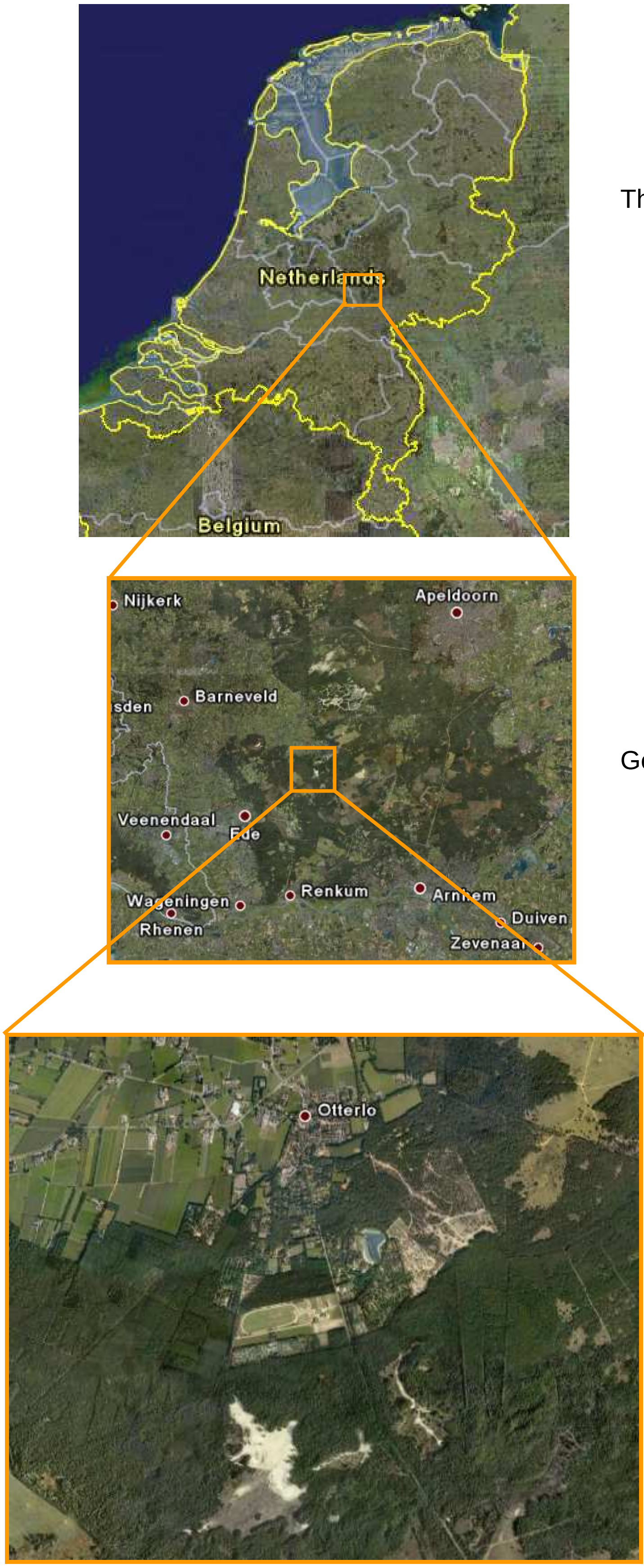
• [url15] http://blog.seniorennet.be/hetpinegeltje2, accessed 21 July 2007

• [url16] http://www.zdnet.fr/actualites/informatique/0,39040745,39370190,00.htm, accessed 24 February 2008

• [url17] http://www.mappy.fr, assessed 12 September 2007

• [url18] http://www.openstreetmap.org, assessed 27 August 2007

Appendix 1: Study area



The Netherlands

Gelderland

Veluwe area around Otterlo

Figure A1.1: Study area.

Appendix 2: New settings for linear regression

Original track from PDA+Bluetooth

Track1 (1 point every second) →

- 1 point every 100 meters
- 1 point every 150 meters
- 1 point every 200 meters
- 1 point every 250 meters
- 1 point every 300 meters
- 1 point every 90 seconds
- 1 point every 120 seconds
- 1 point every 150 seconds
- 1 point every 180 seconds
- 1 point every 210 seconds

Track2 (1 point every second) →

- 1 point every 100 meters
- 1 point every 150 meters
- 1 point every 200 meters
- 1 point every 250 meters
- 1 point every 300 meters
- 1 point every 90 seconds
- 1 point every 120 seconds
- 1 point every 150 seconds
- 1 point every 180 seconds
- 1 point every 210 seconds

Track3 (1 point every second) →

- 1 point every 100 meters
- 1 point every 150 meters
- 1 point every 200 meters
- 1 point every 250 meters
- 1 point every 300 meters
- 1 point every 90 seconds
- 1 point every 120 seconds
- 1 point every 150 seconds
- 1 point every 180 seconds
- 1 point every 210 seconds

Track4 (1 point every second) →

- 1 point every 100 meters
- 1 point every 150 meters
- 1 point every 200 meters
- 1 point every 250 meters
- 1 point every 300 meters
- 1 point every 90 seconds
- 1 point every 120 seconds
- 1 point every 150 seconds
- 1 point every 180 seconds
- 1 point every 210 seconds

Appendix 3: Implementation model in the Veluwe area

• Each data files is reprojected from WGS84 to RD new:

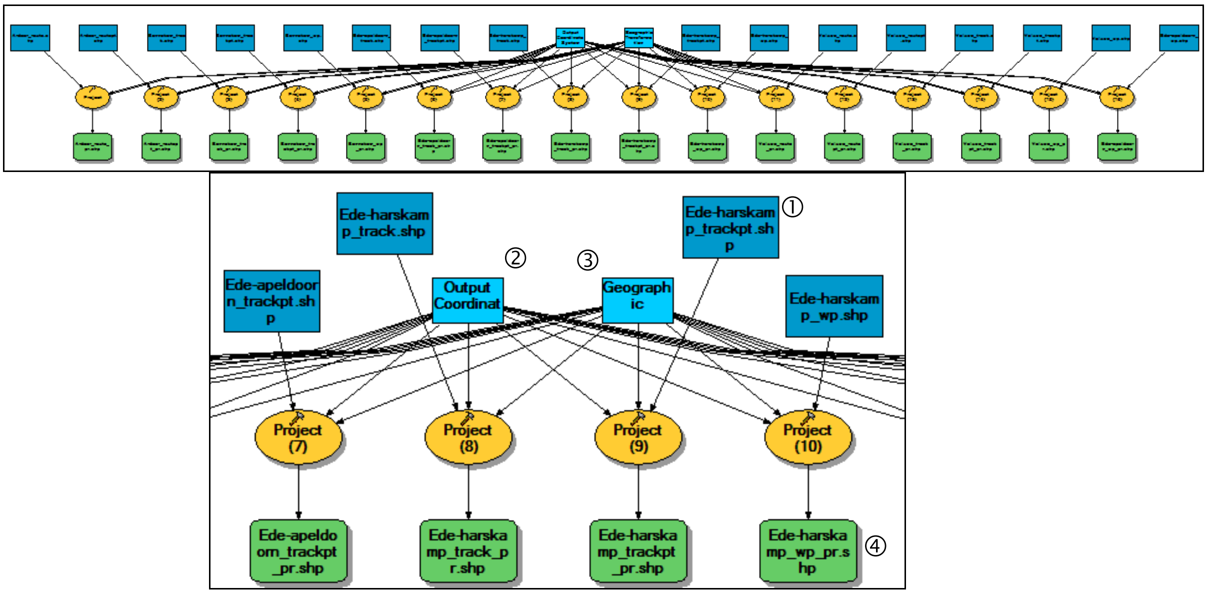


Figure A3.1: Reprojection operation in model builder.

① Original files, that are reprojected according the selection of the suitable
② Projected coordinate system: RD new
③ Geographic transformation from WGS84 to Amersfoort
④ projected files

• Integration and merge

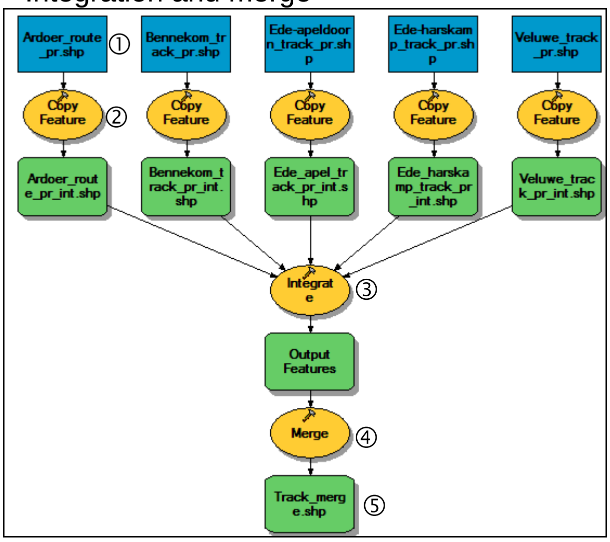


Figure A3.2: Integration operation in model builder.

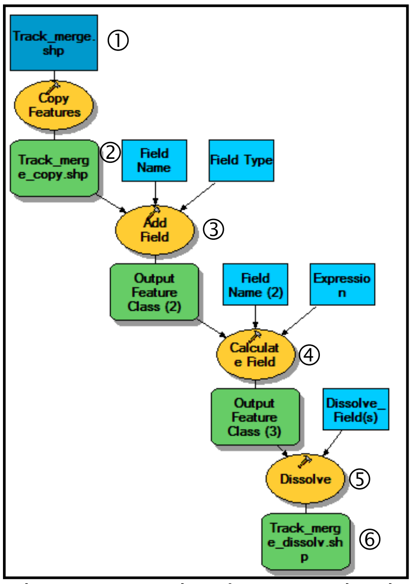


Figure A3.3: Dissolve operation in model builder.

① Projected data
② Copy
③ Integrate with a tolerance of 10 meters
④ Merge
⑤ All data merged in one file

① Merged data
② Copied data
③ Add field called “diss”
④ Calculate field : “diss”=1
⑤ Dissolve
⑥ Final dataset

Appendix 4: Print screen of Georando application

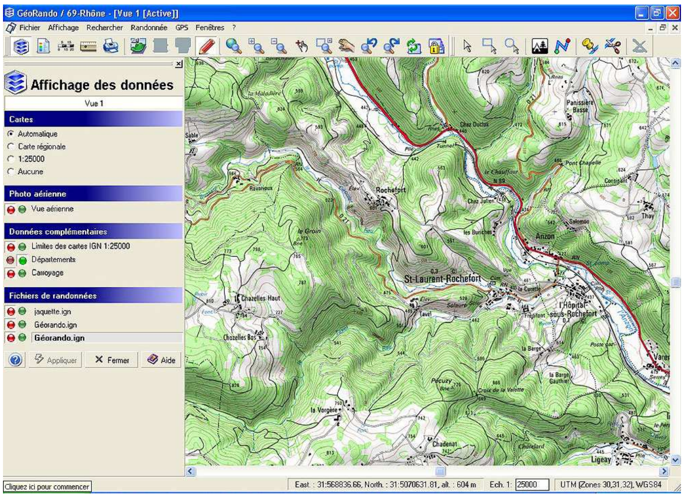


Figure A4.1: 1 : 25 000 based map in Georando.

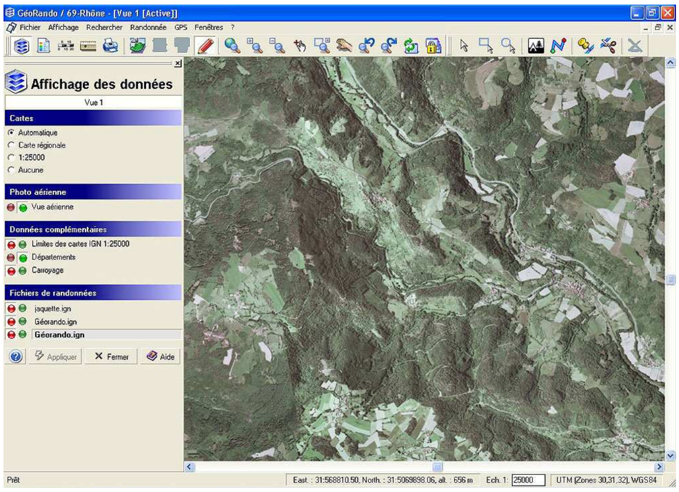


Figure A4.2: Aerial photo in Georando.

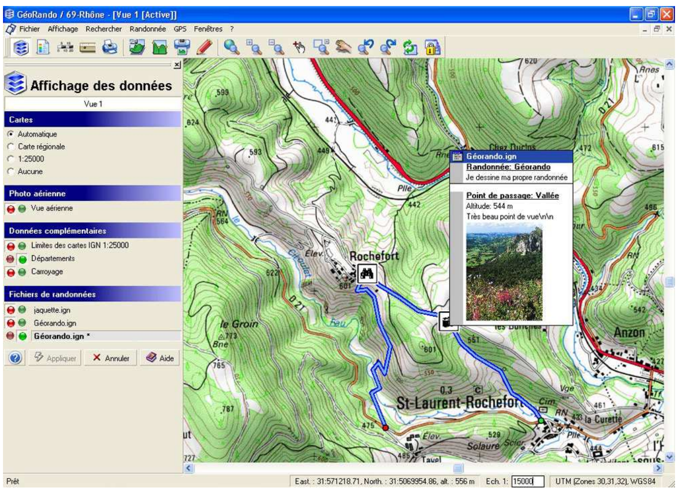


Figure A4.3: 1 : 25 000 based map+ display of route, pictures and comments in Georando.

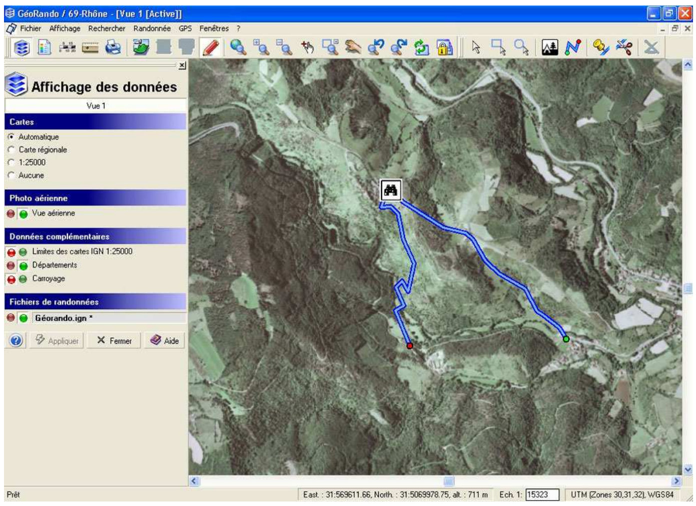


Figure A4.4: Aerial photos+ display of route in Georando.

Appendix 5: Difficulty level integration

• Difficulty level definition

Concerning hiking, the difficulty perception is a individual feeling that can vary according to each person. Then, a difficulty level scale must be defined to suit as close as possible to the perception of the majority of users. This classification has also to fit with the context and the location of the network. Indeed, the classifications find in some mountain books are not relevant for the Dutch landscape.

Most of the time, the difficulty level concerns a particular walk trip and is estimated according the length , the steepness, and the soil quality. The goal is to give an idea to the walkers that don't know the area an idea about the accessibility. Here, the accessibility point is kept as a base of the difficulty level definition but can't be addressed the same way. In fact, this difficulty doesn't concern a particular trip but a part of a network. Then the length can't be considered anymore and the accessibility lay more in the possibility to run into obstacles.

• Difficulty level calculation

The difficulty level defined above can be determined by the combination of the land use and the elevation model information.

The Dutch land use dataset, LGN5, is divide in six mains monitoring classes: agricultural area, forest, water, urban area, infrastructure and nature. According to those classes an "obstacle possibility" value can be defined:

- Infrastructure and urban area classes can be assumed are secure and stable area, where is it easy to travel: easy ➔ 1
- After, there is the agricultural area, that can present some obstacle, like gates or due to natural factors, but it still remind area well-kept by human activity ➔ 2
- The most hazardous area are nature and forest classes where the human maintenance is the less present ➔ 3

The elevation model will be used to calculate the change in altitude. For each segment, the difference between the highest and the lowest point will be calculated. This value will be liked to the total length of the segment in order to calculate a slope percentage. A scale of value will correspond to the importance of the slope:

- If the slope angle is less than 5%, the attributed value is ➔ 1
- If the slope angle is between 5% and 10%, the corresponding value is ➔ 2
- If the slope angle is more than 10%, the value is ➔ 3

A combination of those two information sources will allow to attribute a difficulty level:

Difficulty level = land use + 3 x change in altitude

The change in altitude influence more the difficulty of the track and is a more tangible information. Then this must count more in the difficulty level calculation.

Difficulty scale: from 1 (very easy) to 4 (more difficult)

Land Use	Change in altitude	Difficulty scale	Difficulty description	Corresponding code
1	1	4	Very easy and without possible obstacles	1
	2	7	Very easy and without possible obstacles	1
	3	10	Easy	2
2	1	5	Very easy and rare obstacle	1
	2	8	Easy and rare obstacle	2
	3	11	Medium and rare obstacle	3
3	1	6	Easy and with possible obstacles	2
	2	9	Medium with possible obstacle	3
	3	12	Hard with possible obstacle	4

Table A5.1: Scale of difficulty level.

- **Result**

Because of the study area, the scale defined for the elevation variation has been modified to give a noticeable result. Indeed, the altitude changes were very low, but still allow to compute and test the difficulty level.

The computation of the difficulty level gives the following result:

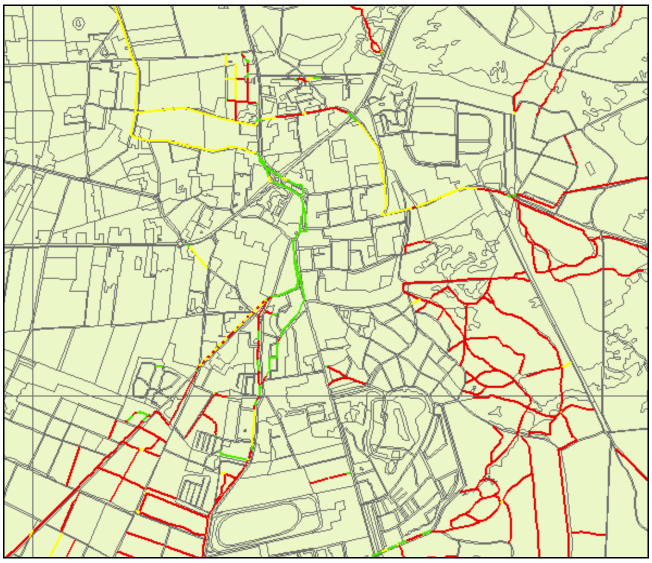


Figure A5.1: Network displayed according the difficulty level, from easy (green) to hard (red).

- **Field test:**

The route travelled for the field application presents different difficulty levels. The test checked the difficulty level, to validate the established one. Each tester has to complete a map showing the route by annotations and signs, to report each time they felled a change in the walk difficulty. The test allowed to check the attributed difficulty level.

The following figures show the correspondence between the difficulty computed and the one established by the testers.

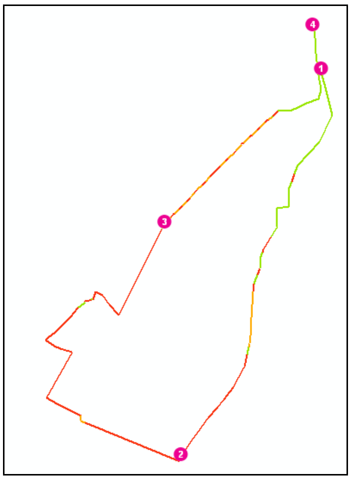


Figure A5.2: Computed difficulty level for the test route.

Figure A5.3: Difficulty level according testers.

The correspondences between the computed and the experimented difficulty level are not really exact. There are similarities between the both, but the tested one is more regular than the computed one which is more chopped. It can be explained by the fact that the computed one is based on raster data. Then the categories are attributed according the cells value and it divides the path according the underneath dataset. Then, it gives this non homogeneous result. Moreover the based map bring its own uncertainty to the result. There is also the point that testers have their own perception of the reality.

In the difficulty level computation, the main part of the value is based on the changes in altitude. This value is calculating by the maximum difference between the highest and the lowest point of the segment, divided by the length of this segment. However, it is a bit restrictive to consider only the maximum change in altitude. The sum of all the positive and negative changes in altitude could be more representative. The same segment with a particular maximum change in altitude won't have the same difficulty if the slope is regular compare to a hilly area where at the end the final change in altitude walked will be four times more. This can be realised by creating the lengthwise profile of each segment and compute the sum of change in altitude.

The validation of this difficulty level attributed to each segment was not really significant because of homogeneous character inherent to the study area. This latter has been selected for the surrounding mixing urban, agricultural and natural area, but didn't present some noticeable variations in height which is the main factor influencing the difficulty level. As the difficulty level computation was the last part of the work, it was not possible to change of study area. So, to balance this and to obtain a result presenting several difficulty level values, the scale corresponding to the differences in altitude had been adapted. So the difference between each value is really low and not really noticeable by the users during the validation in the field.