ESCOLA DE GESTÃO DO PORTO
Universidade do Porto

Miguel Jorge Raposo Nunes de Andrade

A Multi-Criteria approach for the design and evaluation of Demand Responsive Transport (DRT) Services

Escola de Gestão do Porto, Abril de 2008

Miguel Jorge Raposo Nunes de Andrade

A Multi-Criteria approach for the design and evaluation of Demand Responsive Transport (DRT) Services

Orientador:
Professor Dr. Jorge Pinho de Sousa

Dissertação para a obtenção de grau de Mestre em Métodos Quantitativos Aplicados à Gestão

Escola de Gestão do Porto, Abril de 2008

À minha mãe,
aos meus amigos

Quero expressar a minha especial gratidão à EGP, professores e funcionários, por me terem proporcionado um excelente ambiente de aprendizagem e amizade.

Ao meu orientador e professor, Jorge Pinho de Sousa, que esteve sempre presente no decorrer do projecto, mas sobretudo disponível para discutir conceitos ligados com o trabalho e fora dele, discussões que muito valorizo e agradeço, pois são para mim a melhor forma de aprendizagem.

Summary

Demand Responsive Transport services are nowadays an established and recognized way to provide better social inclusion to non-urban, low population density areas. The concept has its origins in the 70's in the USA, but just recently the advances in technology made this type of service more economically attractive. Therefore is of no surprise that this kind of transportation is starting to be a serious option in several regions across Europe.

The European Commission has been promoting projects to study and implement DRT services. One of these projects, MASCARA (de**M**and responsive tr**A**nsport service for increasing **S**ocial **C**ohesion in urb**A**n/**R**ural **A**reas), involved a team from the Engineering School of the University of Porto, with the participation of the author. The MASCARA project was designed to promote the cooperation between several regions in Europe to share knowledge about DRT services, evaluating DRT pilots and DRT feasibility studies.

This dissertation focus on the design of a framework to help this evaluation. The work resulted in combining two known multi-criteria methodologies, namely AHP and TOPSIS. Albeit these methods were designed to help choosing (ranking) between two or more possible scenarios, the proposed framework is able to evaluate and score one scenario situations.

This work can be viewed as having two parts: the first one (from chapter 1 to chapter 4) deals with the design of the framework and methodology. From chapter 5 onwards the practical application is addressed. The author work was focused on the design and software implementation of the evaluation framework.

Sumário

Os chamados Transportes Flexíveis (Demand Responsive Transport), são hoje em dia uma forma estabelecida e reconhecida de combater a exclusão social originada pela falta de transportes em áreas não urbanas, nomeadamente as de baixa densidade populacional. O conceito tem origem nos EUA nos anos 70, mas apenas recentemente, devido aos constantes avanços tecnológicos, é que este tipo de serviço se tornou economicamente atractivo. É portanto sem surpresa que começaram a aparecer projectos de transportes flexíveis um pouco por toda a Europa.

A Comissão Europeia tem também vindo a incentivar projectos para o estudo e implementação de transportes flexíveis. Um destes projectos, MASCARA (de**M**and responsive tr**A**nsport service for increasing **S**ocial **C**ohesion in urb**A**n/**R**ural **A**reas), contou com a participação de uma equipa da Faculdade de Engenharia da Universidade do Porto, da qual autor fez parte. O projecto MASCARA foi pensado para promover a cooperação entre diferentes regiões da Europa, partilhando conhecimento sobre transportes flexíveis, avaliando projectos piloto e estudos de viabilidade de projectos de transportes flexíveis.

O tema desta dissertação é o desenho de uma metodologia para ajudar à avaliação de projectos de viabilidade e projectos piloto de serviços de transportes flexíveis. O trabalho efectuado resultou na combinação de duas conhecidas metodologias de avaliação multi-critério, nomeadamente AHP e TOPSIS. Apesar destas metodologias terem sido criadas para facilitar a escolha entre dois ou mais cenários possíveis, a metodologia desenhada é capaz de avaliar e classificar situações de apenas um cenário.

O trabalho pode ser visto como composto por duas partes: a primeira (do capitulo 1 ao 4) debruça-se sobretudo no desenho e implementação da metodologia. Do capitulo cinco em diante, é dado foco à aplicação prática da metodologia desenhada. O trabalho do autor é principalmente focado no projecto da metodologia de avaliação e na sua implementação em software.

Table of Contents

1 Introduction....1
1.1 Context....1
1.2 Problem definition....2
1.3 The stakeholders....5
Mascara Consortium....5
Individual local (site) teams....5
1.4 General approach....6
2 Demand Responsive Transport Services (DRT) – a brief introduction....9
2.1 Introduction to DRT....9
2.2 Literature on DRT....13
3 Evaluation methodologies considered....16
3.1 Literature review on performance evaluation....16
3.2 Literature review on MCDA....17
3.2.1 AHP....17
3.2.2 ELECTRE....19
3.2.3 PROMETHEE....20
3.2.4 NAIADE....21
3.2.5 TOPSIS....22
3.3 Conclusions and implementation idea....23
4 General evaluation framework....27
4.1 Proposed framework....27
4.2 Practical Implementation....29
4.2.1 AHP....29
4.2.2 TOPSIS....31
4.2.3 TORNADO Plots....32
4.2.4 What-if analysis and monitoring use....33
4.3 Other possible framework extensions....34
4.4 Implementation summary....35
4.4.1 Methodology caveats....35
4.5 Evaluation framework: global view....37
5 Evaluation Plan....38
5.1 MASCARA evaluation approach....38

5.1.1 General principles....38
5.1.2 The evaluation process....39
5.1.3 Global objectives and evaluation model....40
5.1.4 Set-up (phase 1)....41
5.1.5 Data collection and computation (Phase 2)....44
5.2 MASCARA Evaluation procedures and guidelines....48
5.2.1 Objectives and general data requirements....48
5.2.2 Expected results of the evaluation process....48
5.2.3 Relative importance of objectives and criteria....52
5.2.4 Scores....53
5.2.5 The model representation....53
5.2.6 Criteria....54
5.2.7 Objectives....54
5.2.8 Indicators....55
5.2.9 Process phases and data....55
5.3 Data to be collected by each site....59
5.3.1 Site Description....59
5.3.2 Identification of Stakeholders....60
5.3.3 Identification of Expected Impacts and Indicators for Site Objectives....60
5.3.4 Indicator pool....62
5.3.5 Individual Evaluation Plans....62
5.3.6 Evaluation process results....63
6 Evaluation Tool example use....65
6.1 Guided screenshots....65
6.2 Conclusions on the evaluation tool and further uses....74
7 Conclusions and future developments....75

1 Introduction

This dissertation is partially about the author's participation in a European Union Project from April 2006 to December 2007. The main theme of the work is the application of multi-criteria decision analysis and decision support algorithms to Demand Responsive Transport (DRT) services. This type of flexible transport services was the global domain of the project, and the innovation of our contribution comes from developing a new approach for the design and evaluation of DRT.

This chapter introduces the context, the problem, and stakeholders. A graphic representation of the problem approach workflow is presented at the end of the chapter. This representation tries to summarize the work done, main phases and findings.

1.1 Context

INTERREG III is an European Union (EU) program designed to strengthen economic and social cohesion in the EU by promoting cross-border (strand A), trans-national (strand B) and inter-regional (strand C) co-operation.

Partially funded by this program, the MASCARA project (de**M**and responsive tr**A**nsport service for increasing **S**ocial **C**ohesion in urb**A**n/**R**ural **A**reas) in which this work was undertaken, aimed at becoming a vehicle for the promotion and widespread implementation of site-specific collective demand responsive transport services (DRTS) as a key component of social cohesion, sustainable and competitive development in European Regions.

The MASCARA project was a 27-month project (originally planned from 1 April 2005 to 30 June 2007), but later extended to December 2007, under the EU INTERREG IIIC zone West programme and aims to contribute to social cohesion in urban/rural areas through the development of Demand Responsive Transport services (DRT). The project was established to investigate and promote new applications of DRT, which are usually extra to the regular, timetabled Public Transport (PT) services. These services are normally introduced to serve dispersed transportation needs, at off-peak times, between locations not linked by regular PT or where the passengers have more isolated acute needs for transport such as the disabled, the elderly, students, sports/social groups, tourists etc.

In simple terms a typical DRT Service is halfway between a taxi and a regular public transport service. It has to be pre-booked (hours or even several days before the desired departure time) but then it will collect the passengers close to or at their place of departure and take them close to or directly to their destination, picking up other prebooked passengers going in the same direction en route. Often there are different levels of service for different types of customers. For example the elderly or mobility impaired may be collected at the door, whereas more able bodied people may have to go to a nearby stop.

Often the service will promise to collect/drop off people from/to certain points within a certain radius of their location/destination, within a certain time range. A typical guarantee might be "We will collect you within half a kilometer from your home and you won't have to wait for more than 10 minutes."

DRT services can provide valuable, viable services to the community as was previously demonstrated in the following EU projects: SAMPO, SAMPLUS, VIRGIL, SIPTS, INVETE and FAMS. DRT usually work best as part of an inter-modal service chain, linking people with connections to other forms of transport (train stations, expressway bus services etc.).

The MASCARA Consortium included the Faculty of Engineering of the University of Porto (FEUP). FEUP had the leading responsibility for component 4 of the project, the design and implementation of an evaluation plan for the evaluation of the DRT services within the project.
This included the development of an evaluation methodology, plans and support materials. Such evaluation was applied to the Scottish (Angus), Finnish (Tuusla) and Italian (Florence) demonstrations of DRT, and also to the Hungarian (Gyor) and Irish (Cork) feasibility studies of DRT.

The author participated in this project, specially in the design of an evaluation framework and an evaluation tool (software) to support the evaluation of the several MASCARA cases. The project started in 2005, but the author only joined the team in April 2006. Some work had already been done in this component, but was in need of a revision. The commission was to study the problem, and propose changes and enhancements to the original concepts and approach.

1.2 Problem definition

As previously stated, the main goal of this evaluation framework is the evaluation of the individual project implementation of Demand Responsive Transport Services (DRT) in two different contexts: Feasibility projects, and Demonstration projects.

Feasibility projects are essentially focused on the design of the service, considering alternatives within local constraints.

Demonstration projects are focused in validating the service design and finding areas to be improved and/or corrected.

Specifically, the MASCARA consortium was interested in the performance of the services under three main criteria:

- Social Inclusion (SI)
- Sustainable Mobility (SM)
- Technical Performance (TP)

Previous work done in similar DRT projects and the partners experience, suggested a "natural" path to performance evaluation consisting in considering different levels of detail/evaluation, from an aggregated perspective to a higher level of detail. Starting with the three main criteria, objectives were defined within each criteria, then, for each objective, a set of performance indicators was defined to be measured or estimated.

The aggregated values of the set of indicators for each objective will be used to measure of that particular objective performance. Afterwards, the aggregated values for the objectives performance in each criteria will give a measure of the performance for that criterion, and finally, the aggregated values of the three main criteria will yield a global performance result.

This general approach is briefly presented in Figure 1.

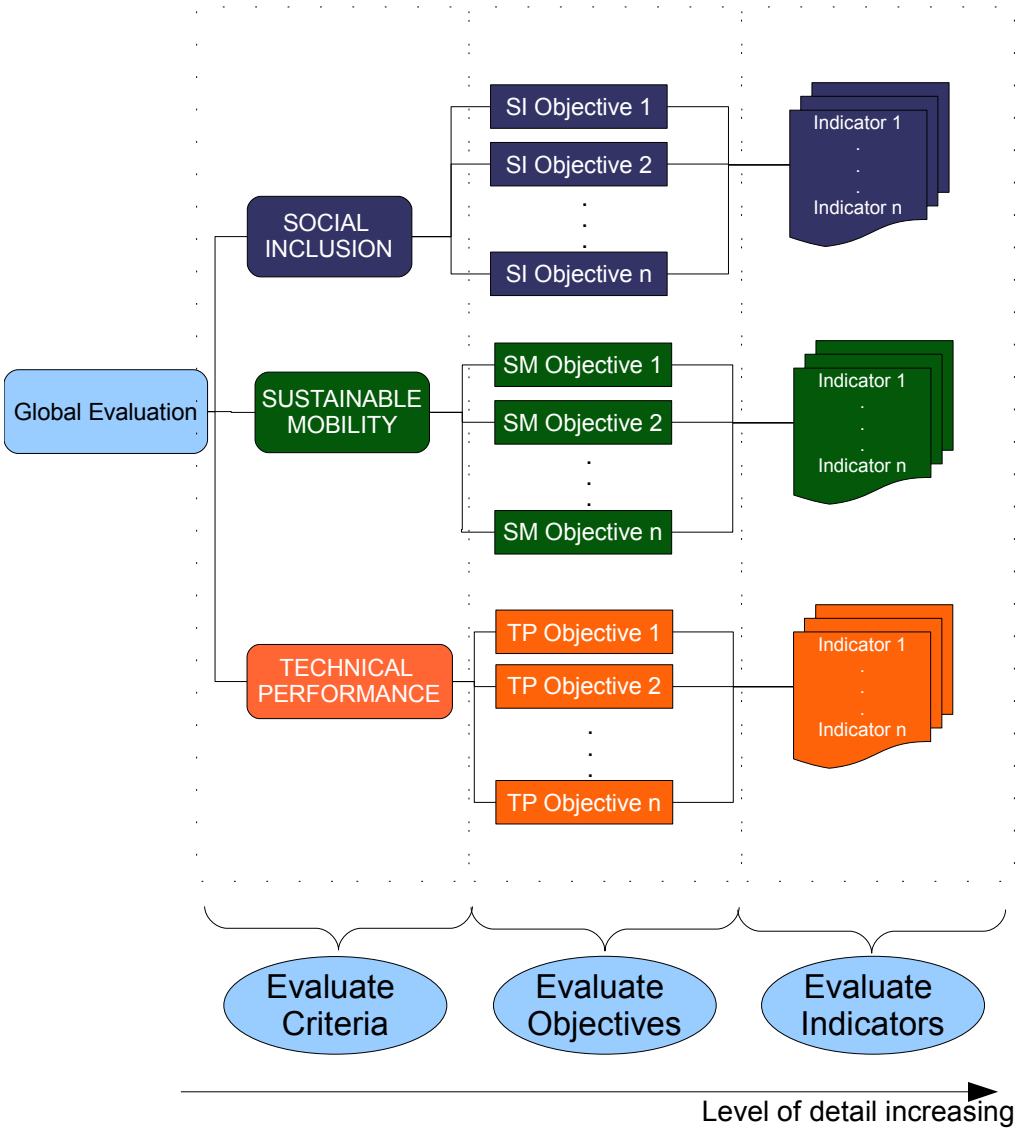


Figure 1: Structure of the evaluation process: For each of the main criteria, objectives are defined. For each objective, a set of performance indicators is defined to be measured

As presented, it is clear that the evaluation process proposed can be viewed as a hierarchical multi-criteria problem, of both quantitative and qualitative nature. Furthermore, the discussions about this approach clearly indicated that for each case the criteria were not equally important, and this relative importance between criteria was different from site to site. The same was found true about objectives within criteria: these are different from site to site, and also differ in importance between themselves.

1.3 The stakeholders

Mascara Consortium

The MASCARA partnership includes 7 Transport Operators / Authorities / Universities from 7 European Regions, representing different transport /mobility environments, ranging from large metropolitan areas to small towns in more rural areas. This provided an ideal context for investigating DRT mobility solutions and achieving useful, tangible, results regarding mobility service accessibility.
The partners involved included:

- Faculdade de Engenharia da Universidade do Porto (Portugal)
- Ghent University (Belgium)
- Central Uusimaa Partnership (Finland)
- Angus Transport Forum (Scotland)
- SITA in Florence (Italy)
- Gyor University (Hungary)
- and the Lead Partner: Cork City Council (Ireland)

Individual local (site) teams

The evaluation process has to be adapted to each particular situation, as the studied cases are different in nature (feasibility or demonstration) and in the service design, which depends heavily on local constraints.

Within each project site, local stakeholders have to be considered. They are divided in three main categories:

- *Political stakeholders* – The DRT services are, to a large extent, dependent on currently adopted policies. DRT schemes often conflict with traditional transport legislation and are in general of a subsidiary nature. It is therefore very important that political stakeholders concerns are addressed in the evaluation.

- *Regional stakeholders* – Transportation has in general a broad range of

influence on several aspects of the region where is operated. These aspects are usually reflected on municipalities and regional organisations attributions.

- *Operational stakeholders* – These are who actually provide the service and know it more intimately, mainly, transport operators.

In the MASCARA consortium, the regional stakeholders were directly represented. It was their job to articulate with political and operational stakeholders, to incorporate their views in the service design.

1.4 General approach

We have two types of evaluation to do: expected performance of newly designed services (feasibility), and actual service implementations (demonstration). Also, the services are very dependant on local constraints, and differ from each other in many aspects. There is only one aspect in common, the "demand responsive" nature of the service, but even here services differ in configuration.

The services to be evaluated are from different regions of the EU with strong socio-cultural differences. The members of the consortium are from quite different cultures and backgrounds, like transports, universities, hired consultancy, municipalities and regional transport organizations.

As referred, the evaluation task is a hierarchical multi-attribute problem, structured in such a way that some guidelines emerged naturally about the design of the evaluation methodology.

The evaluation has a qualitative and quantitative nature. Both aspects must be addressed by the methodology.

This methodology should therefore:

1. be as generic as possible, to be easily adapted to local situations
2. provide mechanisms for evaluation at different levels of detail
3. be made to make it as easy as possible to be understood and applied by a large number of people from different cultures and backgrounds
4. not require elaborate or demanding technology to allow minor practical corrections/changes to be made by local users
5. consider local expert opinions, as opposed to a full automated evaluation

These guidelines are transversal to the whole work, considered in the choice of technology, algorithms, and practical implementations. The same guidelines were also present when designing an *evaluation plan* which is also presented in this work.

Figure 2 summarizes the problem approach and workflow followed in addressing this task, leading to a multi-criteria, multi-attribute evaluation framework. The figure reflects a chronological order of the work in a broad sense. It mentions some methods acronyms that will be clarified later in the text.

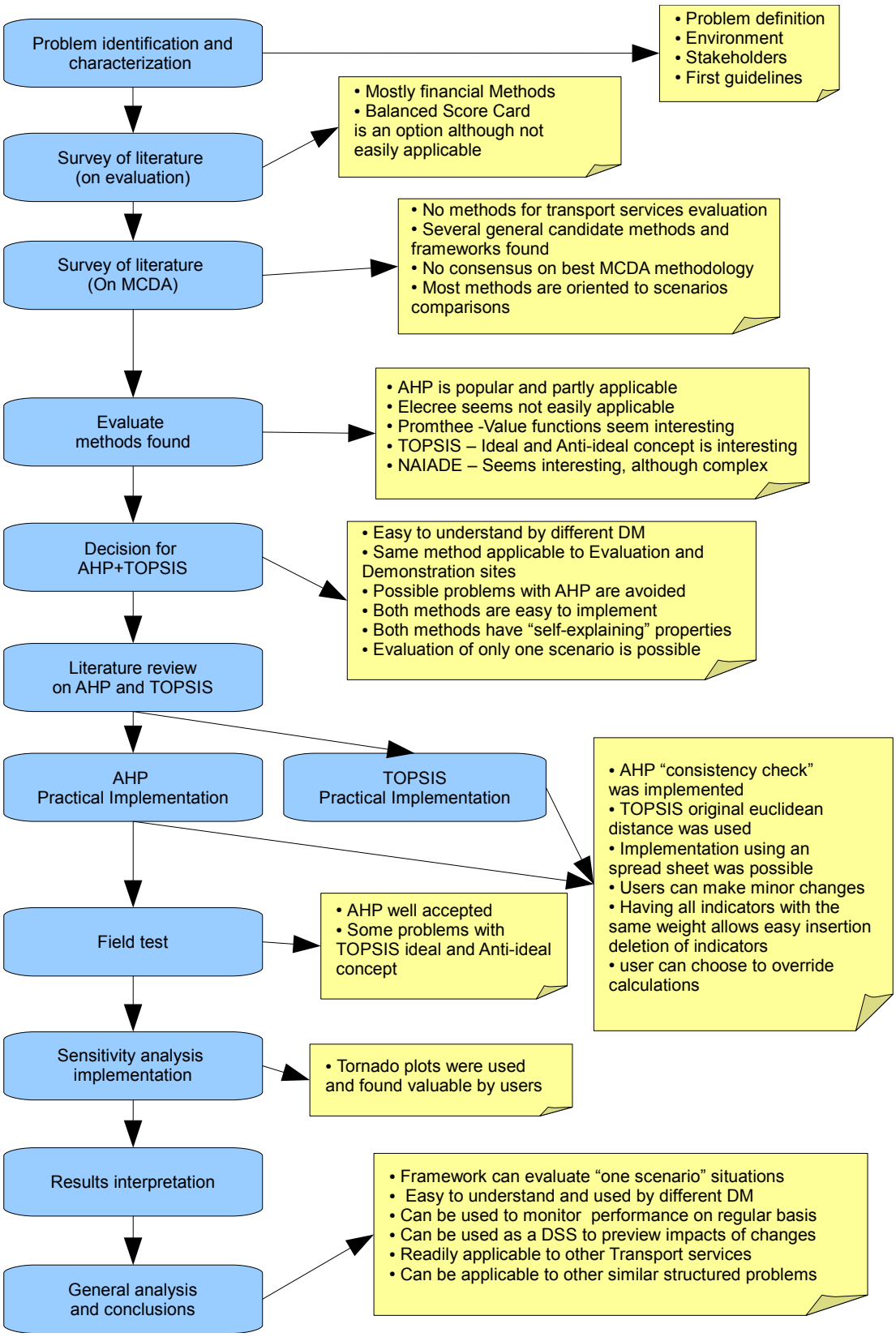


Figure 2: Problem approach workflow (chronological)

2 Demand Responsive Transport Services (DRT) – a brief introduction

In order to design an evaluation methodology, it is helpful to fully understand all aspects concerning the object of the evaluation, in this case, Demand Responsive Transport Services (DRTs). This chapter briefly introduces DRTs [1][2] and presents a literature review on this subject.

No methodology to evaluate DRT services was found on this review.

2.1 Introduction to DRT

Conventional transport (bus services) have fixed routes and fixed timetables, and predefined bus stops. Generally, in the countryside we are faced with large geographical areas with low density population. This means buses have to go to every location that it is meant to serve even when no-one wants to travel. This is obviously sub-optimal in an economical sense, leading sooner or later to low frequency transport services, and decreasing even more the number of passengers due to schedule incompatibilities.

Demand Responsive Transport is a more flexible form of transportation that has been devised to try to match the service offer to the passenger needs.

The taxi is perhaps the simplest form of demand responsive transport. You can call it from wherever you are, and takes you to wherever you want to go. It is convenient but generally is not a widely preferred method of transport, as it is generally associated with a premium price.

From the Demand Responsive concept point of view, we can say that fixed route / fixed timetable (same time every day) is the least demand responsive type of transport, and that a private car is the most demand responsive. The DRT services try to provide bus services with some flexibility.

In a DRT concept, a customer, within a given zone, makes a telephone call to a call centre and requests a single or return trip to go to a nearby town, interchange or local facility. The call centre communicates with the driver of the bus, and the passenger is fitted on to a service round. These rounds may be regular (for example hourly) or entirely according to demand. A pick up point and time are agreed and the bus collects the passenger within a 'time window' of, say 10 minutes, rather like a taxi.

Figures 3 to 10 were adapted form "Good Practice Guide for Demand Responsive Transport Services Using Telematics" [3] and illustrate this kind of service.

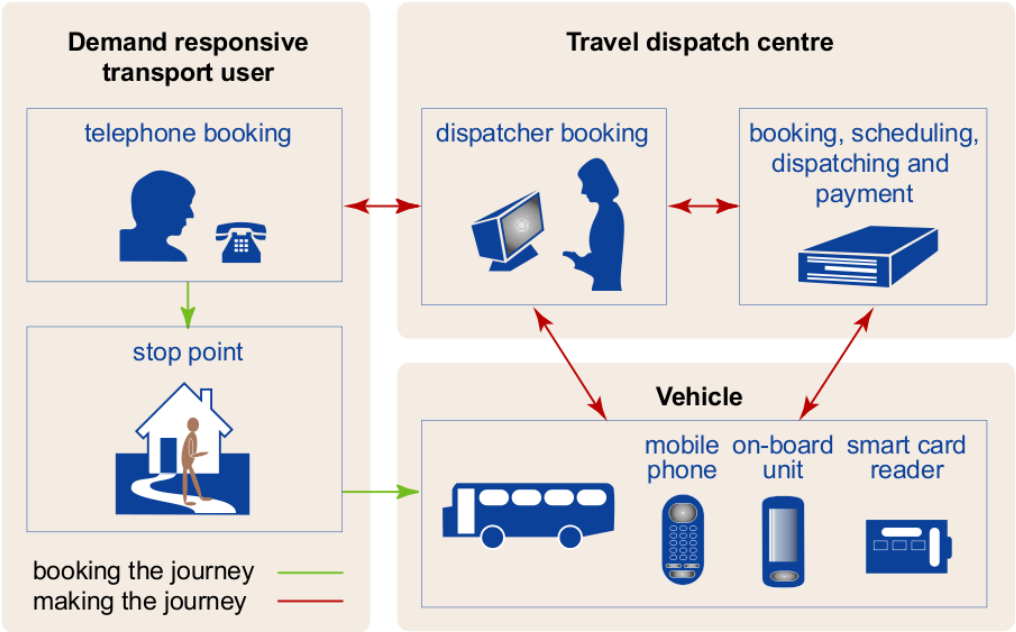


Figure 3: DRT booking scheme

The bus is shared with other passengers with similar requests in a given zone. So, the journey may take longer than a taxi, but the advantages are that the service can be of high quality, almost door to door and be available on a regular basis. Even taxis can be hard to find outside central urban areas.

In addition to this, the computer booking process allows for many more stops to be created, often without any visible sign by the roadside. These do not hold up the bus because it does not have to visit them unless requested. The personal stops can be at or near peoples homes, farms, housing estates or flats and the destinations can be virtually anywhere within a given zone depending on how the scheme is structured. Some buses travel anywhere to anywhere within a zone, while others have some sort of core route from which they deviate.

As well as getting close to door to door travel, booking in itself guarantees better security for the passengers. And, in addition, drivers tend to like the services and enjoy the extra responsibility of seeing that their passengers, many of whom they come to know, are safely returned to home.

Normally DRT services use smaller buses (8 - 25 seats), which are better suited to lower levels of use and can get down narrow, twisting roads and turn round, more easily.

Several DRT route schemes are possible, figures 4 to 10 give an overview of possibilities, following a logic of increasing flexibility:

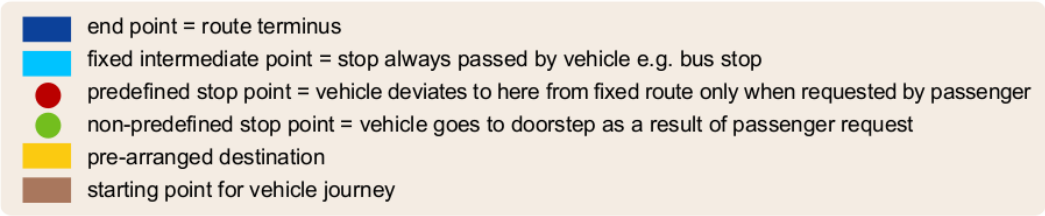


Figure 4: elements for DRT route schemes

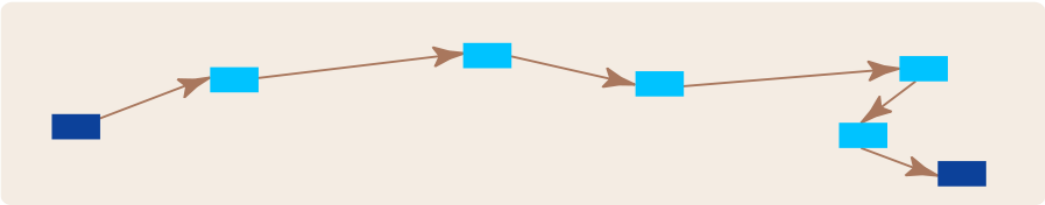


Figure 5: Fixed corridor service (as regular bus but on-demand)

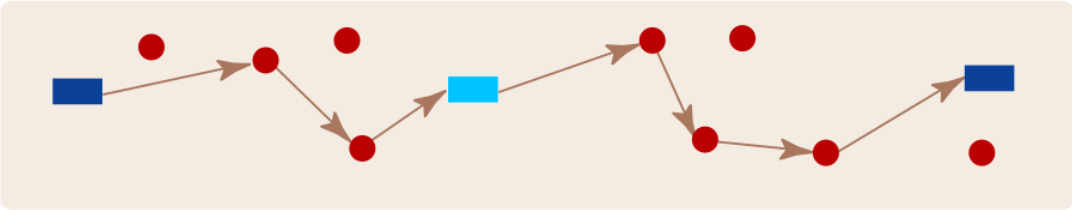


Figure 6: Semi-fixed corridor service

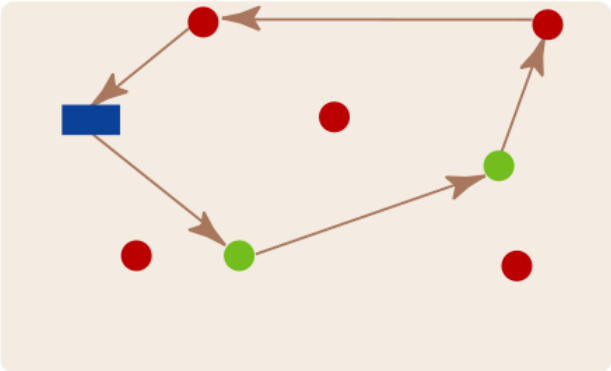


Figure 7: Flexible Area service

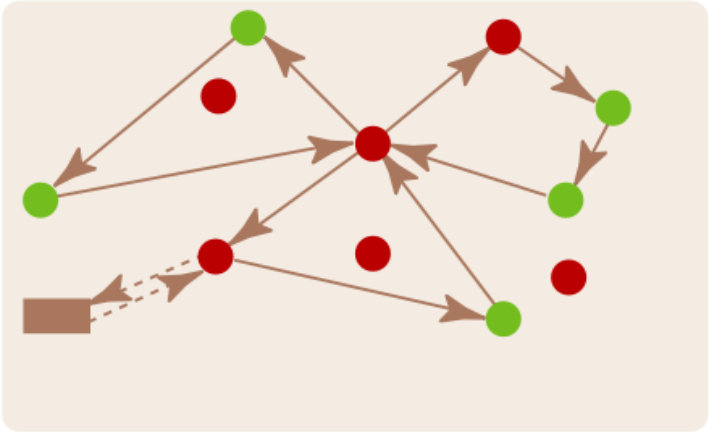


Figure 8: Virtual flexible area service

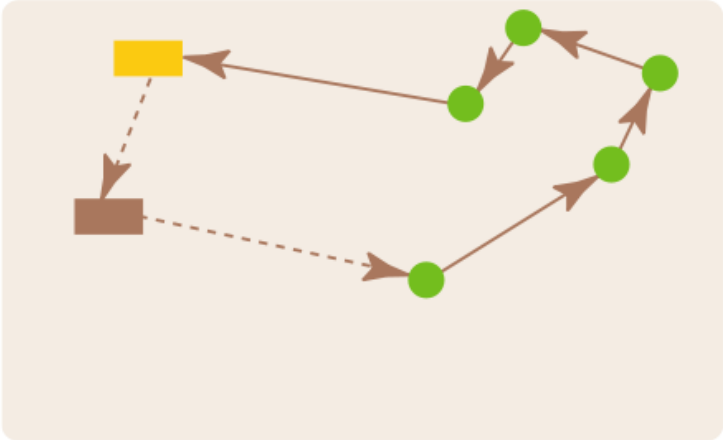


Figure 9: Single destination, multiple pick-ups. (Social care, community transport, Patient transport, Feeder to other main transport services)

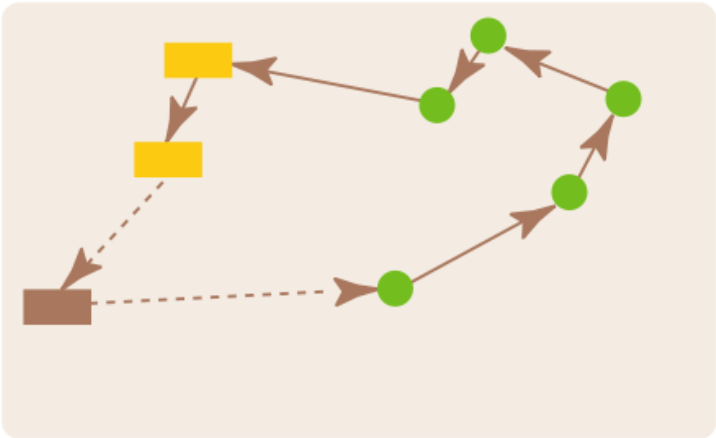


Figure 10: Multiple destination multiple pick-ups, (Education, Work ...)

Other schemes are possible, especially if we consider the articulation with more conventional transportation.

A DRT service has an enormous potential for improved service to passengers, and they are able to serve a wider area with fewer buses. The way this works can be illustrated as follows:

If someone wants to travel from his/her home, in one part of a given zone, but the bus has been booked to go to another, it may not have time to stretch its route and fit them in on that journey. In this case, the call centre offers the passenger the next available bus journey. On the other hand, if the bus is travelling nearby at the right time, the new passenger is simply added to the existing bus round. There is thus an element of negotiation and rescheduling in creating every route the bus will take. This is the key to increased efficiency.

Dispersed demand is not just a rural phenomenon. Other examples are: low density housing in or around urban centres, night travel in urban or rural areas or getting people to and from stations or long distance bus and coach services.

In transport terms, DRT fills a gap between the normal bus service and the taxi. It may still require financial support, but, in the right area, it can offer better service at the same, sometimes lower, cost.

2.2 Literature on DRT

Demand Responsive Transport (DRT) are nowadays established systems to provide transport on low density population areas, low transport demand areas and to population with special needs. Still, the true potential of DRT, as a contributor to sustainable mobility, has not yet been fully exploited – partly because the necessary supporting frameworks required to deliver more effective and competitive transport solutions are not yet fully understood [3].

In the 1970's the diffusion of paratransit services in the USA, (transport services designed mainly to provide health related transportation to people with special needs), started the first designs of DRT. In 1975 Ward [4], proposed different methodologies to model simplified variants of a DRT system in order to compare them with conventional bus line networks.

Since then, several papers have been published on various aspects of DRT systems, mainly focusing design, scheduling and routing problems. More recently, technical advances promoted DRT systems to a new level, making them more efficient and cost effective, targeting a wider range of population, and addressing political and social concerns.

This has lead to a wide increase on DRT services all over Europe, with the EU sponsoring DRT projects since 1998, (Sampos Samplus, Sunrise, Torga, and Mascara).

These programs resulted in a fair amount of DRT implementations (145 sites in Scotland only, [5]) and in a considerable research effort.

Fleet size

The required fleet size for a DRT system to meet a given demand level, with a predetermined level of service, has been studied by Arrillaga and Medville [6] (1974), Flushberg and Wilson [7] (1976), Lerman and Wilson [8] (1974), Daganzo [9] (1978) and more recently Diana et al. [10] (2006) using various methodologies. These works are very focused on costs, and therefore related to the economical aspects of a DTR system albeit their main focus is in technical performance.

Technology

Traditionally, DRT systems need to be a subsidized service. It is imperative for agencies to analyze and investigate their current practices, to identify possible cost reductions or productivity improvements. New advances in technology have been driving studies from several perspectives, like management practices [11], system design [12] and scheduling and route computations, where a large amount of research work has been carried out. Two of the latest comprehensive reviews can be found in Savelsbergh and Sol [13] and in Desaulniers et al. [14]

Multi modal transportation

Additionally some work has been done on integrating DRT systems with other public transport systems, where benefits can be found for passengers, although not easily implemented, since they require passenger comprehension and access to an automatic journey planner, and close synchronization for different kinds of transport, in a multi-agent reality [15][16][17][18][19].

European Projects and research

In Europe, an EU sponsored project (SAMPLUS) dealt directly with technology uses in DRT systems, from which an evaluation methodology for technology impacts was defined [20]. Although they conclude that the potential of DRT systems to reduce transport costs is not fully established, some studies outline the potential positive impacts of Advanced Public Transportation Systems (APTS) on productivity and cost [21][22][23][24][25][26][27].

Furthermore new interests beyond costs / economical are recognized, like social inclusion, quality of life, traffic reduction, pollution reduction, etc. These new interests make DRT systems more appealing [28][29].

Best practice guides

Latest European research, mostly based on experience acquired from EU sponsored programs, has produced some introductory work on aggregating the main body of

DRT knowledge: Halden, D. [30] wrote a guide on How to Plan and Run Flexible and Demand Responsive Transport, a small guide mainly focused on Scotland reality, Brake et al. [3] wrote a "Good practice guide for DRTs using telematics", OECD Publishing published a book "Improving Transport Accessibility for All, Guide to Good Practice" [31] with a chapter dedicated to the implementation of DTR's, and (Halden. D. [32]) published a "Review of Demand Responsive Transport in Scotland" in which common problems and elements of success are reported. These works are important contributions to the design of a comprehensive framework to implement DRT systems.

DRT services evaluation

None of the reviewed papers focus on evaluation methodologies for DRT systems as a whole, although work is in progress in SUNRISE. This project addresses expected and obtained results, evaluated on a set of pre-defined indicators, and the MASCARA project draws on this a step further, implementing a multi-criteria evaluation system, as the basis of a Decision Support System, which is the object of this dissertation.

3 Evaluation methodologies considered

Performance evaluation is a quite general problem, so, a natural step in looking for an applicable methodology was done by performing a literature survey. Previous literature surveys in this field [33][34] were found in the referenced works and were helpful in this task.

3.1 Literature review on performance evaluation

Evaluation of anything implies essentially two steps: some kind of measure of some kind of parameters, followed by some kind of comparison of these measures to other known/expected values.

The evaluation problem, can be easily summarized:

- What to measure (and how)
- What to compare the measure to (and how)

  Generally a derived problem arises from solving these first two questions;

- How to interpret results.

A comprehensive literature survey, lead to the following main findings:

1. Performance evaluation is traditionally achieved by measuring indicators of financial nature [34] or by measuring opinions by use of surveys.
2. The vast majority of performance evaluations is done against predefined objectives, or by comparison with well known results.
3. Recently, evaluation objectives are specified in a form of Key Performance Indicators (KPIs), both of financial, and non-financial natures, and their definition is also a field of management research. Balanced Score Card (BSC) methodology is a popular method when evaluating key performance indicators [33][34][35].
4. Objectives are usually of a quantitative form: e.g. Net return, Number of clients served, percentage of market, etc.

Other methods are mostly from “quality management” literature and generally involve statistical tools and comparison with historical values. Some popular examples are QFD (Quality Function Deployment) methods and Six Sigma.

BSC is about enabling an evaluation strategy definition and finding KPIs to capture the essence of defined strategy. The BSC methodology has been applied to a wide

range of different problems [33][35]. Although it would be acceptable to use the BSC methodology to address our evaluation task, we did not consider it as an option, since no ready available, or easily adaptable framework was found in literature that was applicable to transportation services, and it was not clear how to adapt the BSC methodology to the MASCARA concept. Moreover, the project schedule imposed time restrictions that would make this approach a high risk option.

3.2 Literature review on MCDA

The hierarchical nature of the problem immediately suggested the use of an AHP (Analytical Hierarchical Process) [36] or similar ranking algorithms like ELECTRE [37] or PROMETHEE [38] These are well known algorithms and they were well understood by the FEUP MASCARA team. These algorithms are popular in Multi-Criteria Decision Analysis (MCDA)[1]. These methods are generally used to help choosing between different alternatives and not for the evaluation of a single instance. In order to assure a better understanding of available options, a literature survey was carried out, resulting on several candidate algorithms and some design ideas.

A previous literature survey on MCDA methodologies [39] was helpful in this process. In the next chapter, a summary analysis of each method considered is presented. This is not a comprehensive list of all MCDA methods, only those that we felt appropriate (to an extent) to the task in hands, are addressed.

In what follows, we summarise this literature review on MCDA methods, analysing the pros and cons of each candidate method considered to the design of the evaluation methodology. As referred, this is not a comprehensive review (see [39]) but rather focused on the project objectives.

3.2.1 AHP

General description

The Analytic Hierarchy Process (AHP), is a popular decision making process to help the Decision Maker (DM) set priorities and make the best decision when both tangible and non-tangible aspects of a decision need to be considered. By reducing complex decisions to a series of one-on-one comparisons, then synthesizing the results, AHP not only helps decision makers arrive at the best decision, but also provides a clear rationale of the choice. AHP can efficiently deal with tangible as well as non-tangible attributes, especially where the subjective judgements of different individuals constitute an important part of the decision process. All comparisons are made in a textual form, (e.g. A is slightly better than B) and then converted for calculation proposes in a number, by use of a correspondence scale

[1]Multi-Criteria Decision Making (MCDM) is also a popular acronim for this subject.

consisting if five major degrees (1 3 5 7 9) and four “disambiguation” degrees (2 4 6 8) [40].

Pros

- Most used, well known and generally preferred MCDA method [41][42][43][44]
- Generally easy to use and well understood by DMs [42]
- Probably the most studied MCDA method [41]
- The method generates a descriptive memory of the decision process. That is, for a third party it would be easy to understand the thought process of the DM [40]

Cons

- The method can easily yield a unmanageable number of pair-wise comparisons of attributes and alternatives [40]
- The DM can introduce inconsistency in the analysis [45][46]
- It can surfer from Rank Reversal (RR), i.e. ranking of alternatives can change apparently without reason, e.g. by removing an alternative, the ranking of the remaining alternatives might change. This can be also observed by adding an irrelevant alternative [47]
- Use of arbitrary scales. Empirical research has found “... the perceived meaning of the verbal expressions varies from one subject to the next...” [48]. This means that the process used to compute numerical values is not universal
- A recent paper by Pérez et al (2006) [49] concludes that, priorities changes and also the RR phenomenon (on hierarchies with four or more levels), can also be observed by introducing an indifferent criterion.

Analysis

AHP is a well known method used thoroughly a wide range of practical cases. Its simplicity, easiness of use and understanding by the DM are generally most desirable characteristics. Also, by using verbal comparisons, the method possesses a self explaining characteristic, which we found quite important in our multi-stakeholders environment. On the other hand, it is the “most criticised” MCDA methodology reviewed. For our proposes, some of this criticism is irrelevant, and some avoidable or minimized, albeit it is safe to say that, in a general sense, extra caution is needed when considering this methodology.

3.2.2 ELECTRE

General description

The ELECTRE method was introduced by Benayoun et al in 1966 [50]. The method is based on an outranking concept by using a pairwise comparison of alternatives under each of the appropriate criteria. There are several extensions (ELECTRE IS, II, III-IV, TRI) based on the same principle, but tuned for more specific problems. The DM can indicate indifference about two alternatives under a particular criterion or even that he/she is not able to formulate a judgement. Thus the outranking relations may be complete or incomplete. The method also includes a veto concept, where some (discordant) measure over (or under) a certain threshold can, on its own, exert a veto on an entire outranking relationship.

Pros

- Its a well known method
- It has been successfully applied to a wide range of problems
- It does not require a complete description of preferences
- The concept of out-ranking is easily understandable

Cons

- The method is not multi-attribute by nature, although could be adapted to our case
- Can suffer from rank reversal problem [51]
- Requires that the DM sets preference and indifference thresholds to criteria. This requires that the DM has a good understanding of the criteria and attribute values
- Method is most useful when deciding between many alternatives with few criteria [52] which is not our case
- The method can lead to inconclusive findings about alternatives

Analysis

The outranking nature of this method makes it not readily applicable to our evaluation problem. Our problem has a fair number of criteria (objectives and attributes), and few alternatives, the opposite situation to where this method is best suited.

The adaptation for a multi-attribute application (several attributes concurring for one criteria or objective) seems possible, although not completely clear.

The possibility of use to the evaluation of one scenario, (no alternative scenarios) is not clear.

3.2.3 PROMETHEE

General description

The PROMETHEE [38] methodology for ranking alternatives is based on the same principles as ELECTRE type methods. It provides a partial ranking of the alternatives (focusing on the best alternative) based on preference aggregation. Pairwise alternatives are compared based on (P) preference, (I) indifference or (R) incompatibility. There are also other versions (II, III, IV), PROMETHEE II provides a total preorder of the alternatives. The PROMETHEE method introduces thresholds in the preparatory phase of criteria definition. Some amount of subjectivity is involved, especially in setting preference thresholds, using a set of predefined utility functions.

Pros:

- Both qualitative and quantitative data can be addressed and data can be used in their own units
- It is a well known method and has been successfully applied to many practical problems

Cons:

- The way in which preference information is processed is rather complicated and difficult to explain to non-experts
- Can suffer from rank reversal problem
- Requires a direct assignment of weights to criteria, which can be a difficult process, and can also constitute a method shortcoming [53]

Analysis

For the many aspects of the analysis it is the same as for the ELECTRE method, since both methods have the same conceptual basis.

The concept of value functions on the criteria definition phase is interesting, but it is not a strong requirement in our evaluation task.

As in ELECTRE case, the method does not fit well the hierarchical and multi-attribute nature of our evaluation task, although it could be adapted to this extent.

3.2.4 NAIADE

General description

NAIADE (Novel Approach to Imprecise Assessment and Decision Environments) is a discrete multi-criteria method developed by Munda [54], whose impact (or evaluation) matrix may include either crisp, stochastic or fuzzy measurements of the performance of an alternative with respect to an evaluation criterion.

A peculiarity of NAIADE is the use of conflict analysis procedures to be integrated with the multi-criteria results. NAIADE can give the following information:

• Ranking of the alternatives according to the set of evaluation criteria (i.e. technical compromise solution/s)

• Indications of the distance of the positions of the various interest groups (i.e. possibilities of convergence of interests or coalition formations)

• Ranking of the alternatives according to actors impacts or preferences (social compromise solution).

The whole NAIADE procedure can be divided into four main steps:

• Pairwise comparison of alternatives according to each criterion

• Aggregation of all criteria

• Ranking of alternatives

• Social conflict analysis

Pros

- NAIADE is designed with contributions from several other known methods, and in this sense, is more a framework than a method by itself. It is specially appropriate for the evaluation (assessment) of ill-defined or imprecise scenarios, considering different groups of stakeholders (eventually with conflicting interests). These characteristics seem appropriate to the evaluation of DRT services, especially for feasibility scenarios.
- The method addresses social conflict analysis, which could be important to the design of DRT services and political endorsement. No other methodology was found with this characteristic.
- It uses similar mathematical approaches to other known MCDA methods, and it's fairly easy to understand to MCDA practitioners.
- It was designed to address unique scenarios, where outcomes are fairly unknown.

Cons

- It is a rather new methodology, with a low number of references in literature, mostly the work of the original author, Munda
- It is an elaborate and demanding method, consisting in several interaction steps and analysis with the DM
- Of all methods considered, it is the most technological demanding, although implementation seems feasible.

Analysis

The novelty of this approach is probably the reason for its lack of literature references. It would be hard to implement, due to its complexity, although feasible since it is based on well known methods. It is the only method which specifically addresses social conflict, and this is seen as a general desirable feature on DRT feasibility studies. In the MASCARA feasibility projects, this was not a strong issue mainly due to contingency reasons, but the social conflict aspect of transport service design was recognised by the consortium as an important issue to address.

3.2.5 TOPSIS

General description

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), proposed in 1981 by Hwang [55], is based on the idea that the most preferred alternative should be at the shortest distance from the ideal solution and at the longest distance from the negative ideal solution. This method is quite unique in this approach, although it shares a starting base with ELECTRE methods.

Pros

- The TOPSIS approach, using the concept of ideal solution but also the concept of anti-ideal solution, fits almost perfect our need for evaluating "one scenario" situations. For this, only a conceptual change to the algorithm is necessary, i.e. instead of using the "best of the set" for the ideal value and "worst of the set" for the anti-ideal value, we can use "best we can expect" and "worst we can expect" respectively.
- It uses direct input of attribute values on their own units and scales. This is of great importance when dealing with several attributes of different nature or when there is a need to frequently change, add or remove attributes.
- TOPSIS concept is very simple to explain and easily understood by non-experts [56][57]. It also is very undemanding in calculation needs, and can be easily implemented using relatively simple algebra.

Cons

- Requires direct input for attribute weights. This is generally against field best practices of not asking the weights to the decision manager in a direct manner [53].
- It assumes all attribute values have a monotonic behaviour between the indicator value and it's goodness, e.g. the higher the better (or worst). Situations where this cannot be assumed can be easily foreseen, and special treatment would become necessary.
- The method by it self does not provide any form of hierarchical evaluation.
- TOPSIS can also suffer from the rank reversal problem [47].

Analysis

The TOPSIS was the only considered method that was somewhat readily applicable to our "one scenario" evaluation requirement. It has some good characteristics of great value to our case, in particular the intuitive nature of the method become highly important, when targeting a use by a broad range of stakeholders, to which the methodology must be explained. The method does not fit the hierarchical aspects of our evaluation necessity.

3.3 Conclusions and implementation idea

A common aspect of all the MCDA methods reviewed, is that are all orientated to choosing (ranking) from different scenarios. This immediately poses a problem to our evaluation task, that is, how to evaluate a one scenario situation?

A first idea was to construct a fictional "average" scenario and try to compare the proposed (expected) service against this one. In the same line of thought, it could be easier to construct a fictional "ideal" scenario, based on best possible results, and compare the (expected) proposed scenario with this one.

Even if the methodology provided a numerical evaluation of the proposed scenario against the fictional one (ideal or average), it would always be relative. For example, one could find that some scenario was far from ideal, but what does that distance mean? Is it a very bad, simply bad or an acceptable situation?

The TOPSIS approach

The TOPSIS methodology shed some light on this matter. The method uses the concept of ideal and anti-ideal solutions. In general, the "ideal" is chosen to be the best of a set of alternatives, and the "anti-ideal" the worst of the same set of alternatives. It is not difficult to apply this concept with a minor change: instead of real alternatives, we can use two fictional scenarios, one ideal and one anti-ideal. So, for a particular indicator (attribute) of a feasibility (or actual) scenario, one should

ask what is the very best and the very worst it can do.

Using this process, we have a way to aggregate a set of indicator values into an objective score. Figure 11 illustrates the TOPSIS scoring of a set of indicators.

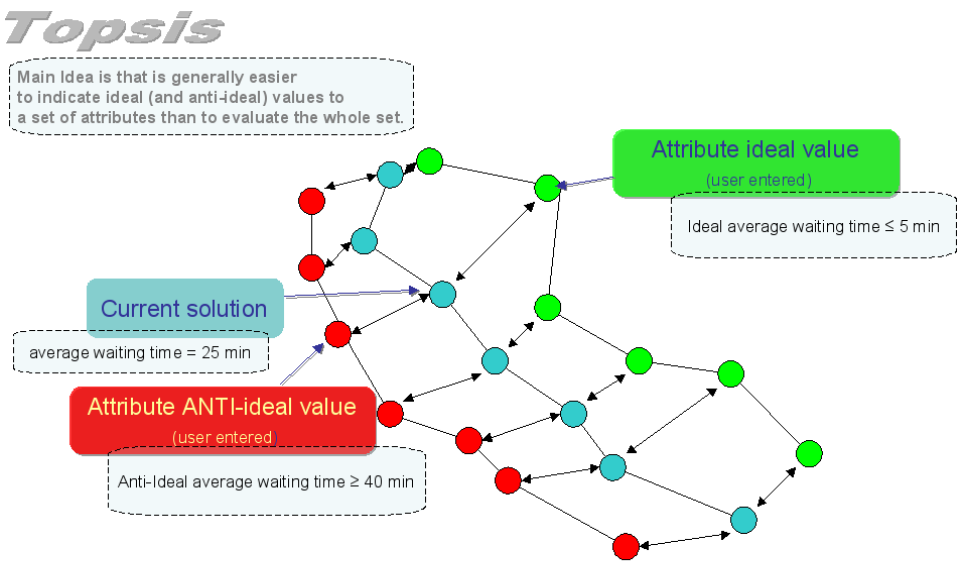


Figure 11: TOPSIS illustration. For a set of (numerical) indicators the method can compute a score based on the distance of the current solution to the ideal and anti-ideal hypotheses

For each indicator, current, ideal and anti-ideal values can be expressed in their natural units, (e.g. min., km, Euro, etc.) which are normalised later. This largely facilitates the data input. For the group (set) of indicators, The TOPSIS algorithm computes a score based on the group distance to the "anti-ideal" scenario (farther the better) and on the group distance to the "ideal" scenario (closest the better).

If we use a textual rating scale (like AHP) with a numerical correspondence, (i.e. very bad – 1 ;bad – 2; acceptable – 3 , good – 4; very good – 5) where the score 1 would be assigned to the "anti-ideal" scenario and 5 to the "ideal" scenario, it would be straightforward to compute a score (1-5, absolute ranking position) for that particular set of indicators.

By using this simple TOPSIS approach, we have a way to compute an objective score. We now need a way to repeat this aggregation process for the criteria, and for the global score.

This is a somewhat trivial task, achieved by multiplying the previously found objective score by its importance weight within the criterion.

Example

Criterion "Technical Performance" with further specification of objectives:

Technical performance		
Objectives	Weight %	Score [1..5]
System speed	15	3
System failure	20	4
System accuracy/ reliability	15	3
System acceptance	30	2
Telecommunications	20	5

Table 1: Objective weights and scores (scores found by TOPSIS) for a specific criterion example

In this example, the "Technical Performance" criterion score would be:

$$0.15\times3+0.2\times4+0.15\times3+0.3\times2+0.2\times5=3.3$$

The same logic can be used to aggregate all the criteria scores into one global score.

We have now an idea for a method to address our evaluation problem: we ask the DM (who could use the opinions of technical people) to find ideal and anti-ideal values for each indicator to be measured, specifying that "ideal" means the "best we can hope for" and "anti-ideal" "worst we could do" for the particular situation. This will allow us to evaluate one or more scenarios, based on the same ideal and anti-ideal fictional scenarios.

AHP approach

We can now concentrate on finding the weights for the objectives and criteria. Of course, these weights could be asked directly, but the AHP method already provides a simple way of finding criterion weights, and it can be used in a hierarchical multi-level way. Since we are proposing the AHP method to find criterion weights only, most of the methodology caveats are avoided.

This particular use of the AHP is not new, and there are several references in the literature to this application including in combination with the TOPSIS method [40][42][58][59][60][56][61][62].

In the literature we found that the AHP is mostly used to find the weights to TOPSIS criteria, on a one level (non-hierarchical) situation. Our proposed implementation differs slightly: we handle all indicators as having the same weight, thus allowing easy addition or removal of individual indicators from a set. This was decided based on practical reasons, as the indicator sets are most unstable and removing / inserting / changing indicators is frequent. Also, the indicator sets can be reasonable large, which might be a problem to evaluate all pairwise comparisons needed by the AHP method.

These are the main ideas used for the design concept of the evaluation methodology. In the next chapter we will present the framework we have developed in more detail and focus on practical aspects of the tool development and its use.

4 General evaluation framework

4.1 Proposed framework

The evaluation framework presented in this dissertation, consists essentially in two steps. In a first step, a conceptual work is done, by specifying clear objectives for each criterion. These objectives should reflect each site specific situation, and they are expected to be different from site to site, although knowledge of other location objectives could be valuable and used.

First step - finding criteria and objective weights

In this phase, The AHP method is used to measure criteria and objectives relative importance resulting in a weighted sum. We use a five point scale for pairwise comparisons between criteria (and objectives) namely,

A is “Equally important” than B
A is “Slightly moderately more important” than B
A is “Strongly more important” than B
A is “Very strongly more important” than B
A is “Extremely more important” than B

Table 2: AHP textual comparisons

For each textual comparison, the Saaty [36] correspondence scale is used, respectively (1 3 5 7 9). The correspondence is done automatically and is not presented to the user. We feel that keeping a textual description has an important informative value, which could be lost if the user was asked directly for the correspondence.

Our methodology deviates from Saaty's slightly as we chose not to present the user (DM) with the disambiguation scale (corresponding to values 2, 4 6 and 8). The main reason to do this was to follow the guidelines (see section 1.4) and keep the methodology as simple as possible. It would add unnecessary complications and the necessity for “fine tuning” in this area was not felt in the field. The disambiguation scale could always be inserted if needed at some point. Research in other fields also indicate that a five level scale is a “good” safe number of levels when considering several different stakeholders [63].

By using AHP only to find criteria and objective weights, we avoid several of the

methodology known problems, but still some remain of concern. Since we are working with stakeholders from different cultures and backgrounds, the textual scale used can introduce some differences of interpretation [48], so we opted to show the calculated weights as a result of the AHP process. The DM is then given a chance to review, accept or change the weights directly.

Second step - evaluation of a set of indicators

In a second step of the evaluation process, sets of (measurable) indicators are defined for each objective. Then, for each indicator, ideal and anti-ideal values are entered. This allows the TOPSIS method to compute a score for each set of indicators (objective score), which are then aggregated into criteria scores and finally into a global evaluation score.

Due to the different nature of the projects we need to evaluate, a special attention to the definition of indicators is needed. For the feasibility cases, sometimes there was no clear idea of what value to expect from an indicator. A classification of indicators was designed to capture this information. An indicator could be:

- M – Measured
- EL – Estimated based on Local knowledge
- EO – Estimated based on Other sites experience
- G – Set as Goal.

The indicator could also assume a textual form, in case there was a need to "contextualise" a set. These textual indicators obviously cannot be considered on computations. The user is also given the opportunity of choosing which indicators will be used in TOPSIS algorithm, so that less confidence indicators could be left out if desirable.

Due to the possibility of a fair amount of uncertainty about indicator values, the framework uses the TOPSIS only as an aid. The computed score is shown to the DM and then he/she has a chance to accept it by entering the same score on a side-by-side column, or correct it by entering a new score. In case of correction, a short explanation is expected. In any case, if the expert evaluation differs substantially from the calculated value, it would most certainly indicate a judgement error, or that the particular set of indicators is ill-defined for that particular objective.

After the computation of scores, an interactive process can begin by performing a sensitivity analysis on the data, which can be semi-automatically based on predefined Tornado plots (see section 4.2.3 page 32), and by performing "what-if" evaluations. In an extended use, the evaluation tool can be further used as a monitoring tool and/or a decision support system.

4.2 Practical Implementation

4.2.1 AHP

As previously mentioned, we chose to use textual comparisons directly. This was achieved by constructing an interface where, by the use of a “drop-down” widget, the user could input all the necessary comparisons. This is illustrated in Figures 12 and 13:

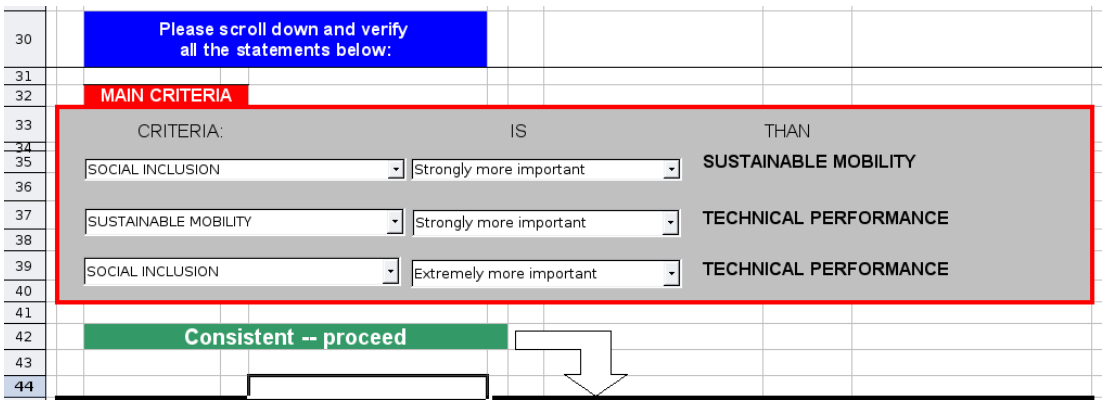


Figure 12: AHP necessary comparisons. User interface

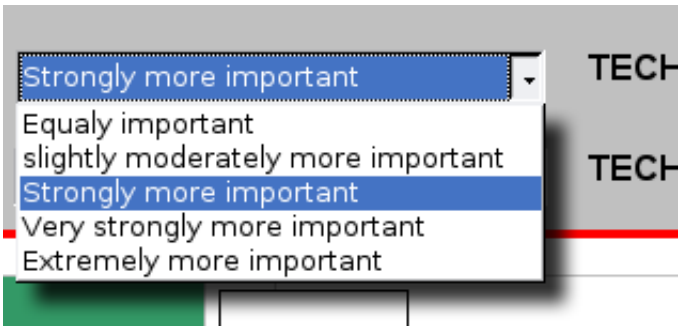


Figure 13: Levels are chosen by a "drop-down" widget. Criteria can be reversed by choosing the alternative in the first drop down

For each possible pair, the user selects which comparisons better describe the situation. In the first “drop down” the user can reverse the two criteria, e.g. for the first comparison, the user can chose “Sustainable Mobility” on the first drop down, and the second criterion would change to “Social inclusion”

Based on these choices, an AHP matrix is evaluated instantly. This is done without user intervention, but can be checked. (See Figure 14).

	Criteria		1	2	3		Weight	
			SOCIAL INCLUSION	SUSTAINABLE MOBILITY	TECHNICAL PERFORMANCE		%	
1	SOCIAL INCLUSION		1	3	5	►	63%	15.00
2	SUSTAINABLE MOBILITY		1/3	1	3	►	26%	1.00
3	TECHNICAL PERFORMANCE		1/5	1/3	1	►	11%	0.07
			1 5/9	4 1/3	9		100%	16.07
	Consistent -- proceed							

Figure 14: Computation of weights. Usually not seen by users.

Consistency ratio

Also, a consistency value is computed and immediately shown. This is to help avoiding inconsistencies in the user choices, e.g. A>B, B>C, C>A. Values shown can be "Consistent - proceed", "Marginally consistent" and "Inconsistent – Review". The AHP methodology allows some degree of inconsistency in user choices. The computation for this occurs in the same sheet previously shown. Figure 15 illustrates this.

		1.95		3.05						
	Aw	0.79	Aw/w	3.06						
		0.32		3.05						
			average	3.06	λmax					
	CI (consistency index)			0.03		(λmax - n) / (n - 1)				
	RI (random const. index (3))			0.52	Table					
	CR (consistency ratio)			5.31%	CI/RI		Consistent -- proceed			
Table										
	n	1	3	4	5	6	7	8	9	10
	RI	0	0.52	0.89	1.11	1.25	1.35	1.4	1.45	1.49

Figure 15: Consistency ratio computation. Not usually seen by users

Practical implementation

The AHP method calls for the computation of the matrix eigenvector. We used a known approximation to this, by first normalising each column (i.e. divide each entry by its column total) and then computing the average values of row i in the normalised matrix which is the estimate for the weight I.

For the consistency computation we used Saaty proposed method using a reference table for random index [64].

4.2.2 TOPSIS

The data for the TOPSIS computation was entered in a sheet dedicated to objectives and indicators. Figure 16 and 17 illustrate the interface.

Sub-Criteria	Weight	SCORE	AID		Support Indicators Description	Value of Indicator	D,M,EL,E O,G	Measurement units
	%							
	17%	4	2	1.1.1	service description		D	text
				new	operational time period of DRT		D	text
				1.1.3	Number of stop points	20	EL	#
				1.4.1.	% of resident population served in area	10	EL	%
Relevant and attractive service				1.4.2.	Characteristics of target population	2	D	text
				1.4.5	Accessibility of passengers to vehicles (percentage of passengers who have problems using public transport according to personal mobility and space? constraints)	3		%
				1.7.4	impact on social inclusion (number of new trips generated by DRT and not the transition from other means of transport)	10	EL	#

Figure 16: Indicator descriptions and values. Columns continue to next figure

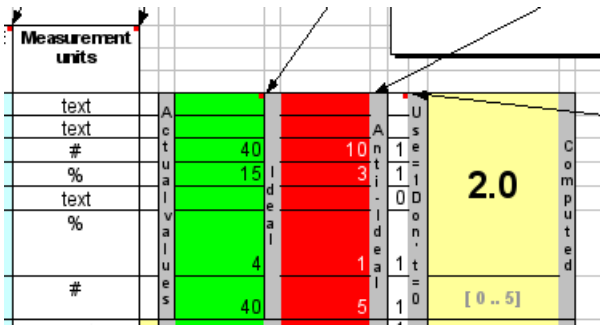


Figure 17: Ideal, anti-ideal values and TOPSIS computation (continued)

In these figures some of the indicators are of "text" kind (text not shown). These are used to better describe a particular situation. The computed value (2.0) can only take into consideration the numerical indicators. As stated, this is just an aid. The user must accept this value explicitly or enter a different one and, in this last case, provide some explanation.

The TOPSIS computation occurs automatically without user intervention, but, as in the AHP case, the computation process can bee inspected in another sheet.

The TOPSIS calculations are straightforward and easily accomplished using standard spreadsheet functions [52]. These calculations are represented in figure 18.

# indicators	7							
ROW	**1**	2	3	4	5	6		
OBJECTIVE S1	Criteria 1 score	Criteria 2 score	Criteria 3 score	Criteria 5 score	Criteria 6 score	Criteria 7 score		
Valor actual	0	0	20	10	0	3		
Ideal	0	0	40	15	0	4		
anti-ideal	0	0	10	3	0	1		
Gij^2	0	0	400	100	0	9		
Gij^2	0	0	1600	225	0	16		
Gij^2	0	0	100	9	0	1		
Sum Gij^2	**0**	**0**	**2100**	**334**	**0**	**26**		
	#DIV/0!	#DIV/0!	0.4364	0.5472	#DIV/0!	0.5883		
	#DIV/0!	#DIV/0!	0.8729	0.8208	#DIV/0!	0.7845		
	#DIV/0!	#DIV/0!	0.2182	0.1642	#DIV/0!	0.1961		
	1	2	3	4	5	6		
	1	1	1	1	1	1	SUM	
Pesos	0	0	1	1	0	1	4	
valor pesado	#DIV/0!	#DIV/0!	0.4364	0.5472	#DIV/0!	0.5883		
ideal pesado	#DIV/0!	#DIV/0!	0.8729	0.8208	#DIV/0!	0.7845		
anti-ideal pesado	#DIV/0!	#DIV/0!	0.2182	0.1642	#DIV/0!	0.1961		
MAX (ideal vs anti)	#DIV/0!	#DIV/0!	0.8729	0.8208	#DIV/0!	0.7845		
MIN (ideal vs anti)	#DIV/0!	#DIV/0!	0.2182	0.1642	#DIV/0!	0.1961		
Desvio ao ideal	#DIV/0!	#DIV/0!	-0.4364	-0.2736	#DIV/0!	-0.1961		
Desvio ao anti-ideal	#DIV/0!	#DIV/0!	0.218	0.383	#DIV/0!	0.392		
								D+ao ideal
Desvio ao ideal^2	0.000	0.000	0.190	0.075	0.000	0.038		0.9086
								D- ao anti-ideal
esvio ao anti-ideal^2	0.000	0.000	0.048	0.147	0.000	0.154		0.6022
	score	**0.398607**						

Figure 18: TOPSIS calculations. Normally not seen by the user

In this figure the computed value is 0.398607 which is in a [0,1] scale. When multiplied by 5 and rounded to the first decimal place, we get 2.0. Also note that text indicators produce some errors (#DIV/0! warnings), these are detected using an error finding function (ISERR) and not taken into calculations.

4.2.3 TORNADO Plots

Tornado plots were at first introduced as a methodology add-on possibility, focusing on sensitivity analysis. It was soon recognised that this was trivial to obtain and of great informative value. We have then decided to include these plots as part of a technical analysis report on each site individual set-ups.

Figure 19 shows a real application example. In this case, the impact on the global score by variations to objective scores is analysed. The plot presents the objectives ordered by sensitivity degree (most sensitive at top, least at bottom). The plot is constructed by varying one objective within the allowed range of variation keeping all other values constant, and then measuring the impact on a dependant function. For the three topmost objectives, we can see that each one is capable of varying the global score from approx 3.2 to 3.9. The fourth objective (attractiveness for families and visitors) is capable on its own to raise global score above 4 to approx. 4.1.

Other type of plots possible, e.g. measuring the solution robustness to the weights assigned to each objective, impacts of indicators on criteria, etc.

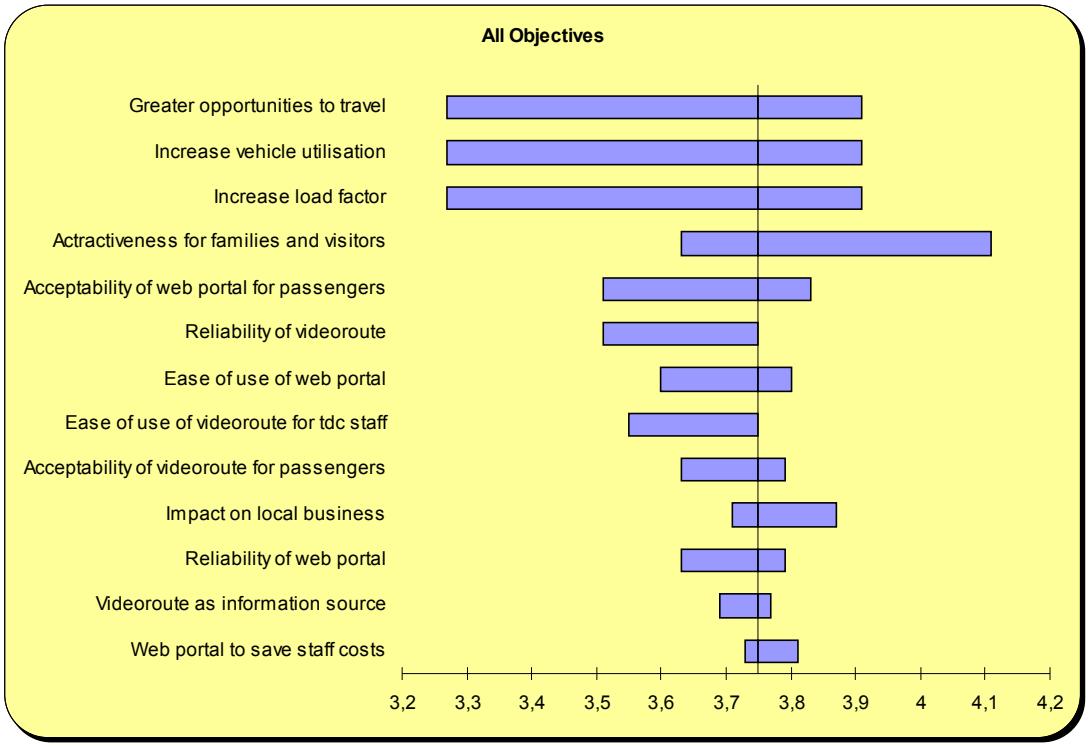


Figure 19: Example Tornado Plot. Shows how the final score would change by varying objective values. The chart orders from top to bottom the most sensitive objective to the least sensitive.

Tornado plots provide a very intuitive picture of sensitivity analysis, without the need of elaborated explanations, which make their use a valuable management decision tool. First we considered doing sensitivity analysis "in-house" and only publishing conclusions on a technical report, but the value of this kind of representations and the fact they could be easily implemented on an software spreadsheet, lead us to incorporate them on the evaluation tool in predefined forms. Note that although informative, Tornado plots by themselves do avoid dispense a closer look at the data. In fact, the technical/sensitivity analysis done for each site revealed interesting aspects that could not be found by a more "automated" procedure, e.g. the use of the same indicator(s) for more than one objective, would amplify the changes effect, and this should be considered as it is not reflected on the plots, which are drawn for one objective at a time.

Tornado plots are quite popular, and there are freely available on the Internet several spreadsheets and spreadsheet *add-ins* that implement the necessary functionality to produce these graphics. We can also find general instructions on how to implement this functionality ourselves [65] .

4.2.4 What-if analysis and monitoring use

The evaluation tool implemented in framework can be readily used to perform "what-if" analysis. The DM formulates an hypothesis, and find what would change

in terms of indicators (or criteria/objective weights) and proceeds to remove/insert indicators, and/or correct indicator values. The global/criteria/objective scores for each change are calculated immediately.

Another natural use for the tool is to monitor a particular service implementation, by measuring the chosen indicators at regular time intervals. A simpler use could be just to measure the most sensible indicators. Collecting data at regular intervals would also provide statistical information that could be used in planning or reacting to changes in the service.

4.3 Other possible framework extensions

The framework development and implementation, led to several ideas to improve the evaluation tool. Some of these ideas were abandoned in light of the framework initial guidelines, mainly conceding to simplicity of use and understanding. Yet some ideas were found worth mentioning as extensions to the tool. One of them, Tornado plots on sensitivity analysis, was actually implemented.

Indicator weights and value functions

By design, the framework does not take into account different levels of importance at the indicator level. This has the immediate advantage of allowing introducing, changing or removing indicators without any further work. This was found a desirable feature during design and implementation. Yet, it is possible that some particular new situation requires differential weighting on indicators, or that this necessity is found in a fine-tuning future phase, when a particular service is stabilized. The TOPSIS method already considers weighting, so this could be accomplished by direct introduction the indicator weighs, or using some other method to find them.

The TOPSIS method assumes indicators to have a monotonic behaviour. This might not be desirable on some situations. For example, consider the “number of calls to book a trip” a high or an increasing value for this indicator would be considered positive, yet, for some value threshold, it would mean that the system would be overloaded, and an increase would mean that the service could be actually getting worse. Most of these situations could be avoided by carefully choosing the indicators, but it would be interesting to have the possibility of recurring to value functions, e.g. as used by the PROMETHEE methodology.

Veto capability

This would be straightforward to implement, but no real interest was shown by any partners on this functionality. The main idea is the same of the one present to the ELECTRE methodology. In our case, it could be applied to any evaluation phase. If, for example, the “passenger pick-up lateness” indicator were above 20 minutes, the

corresponding objective “Customer service” would be scored as “Bad” even if all other of the remaining indicators had top values.

NAIADE

The NAIADE method provides social conflict analysis capabilities. This would be an interesting addition to the framework. On DRT services implementation there are often conflicting stakeholders, such as other transport operators, non-served population, etc. The DRT have also, in the present, a heavy political component. NAIADE seems promising in addressing issues of this nature. Most of the work done by the proposed framework could be easily reused in the NAIADE approach, and thus providing additional useful information, especially for feasibility studies.

4.4 Implementation summary

After the evaluation of the methodologies available, a decision for a framework resulted in a choice of a mix solution using AHP and TOPSIS. Main advantages seen were:

AHP and TOPSIS are very well known methods and generally well accepted and understood by the DM.

AHP has several caveats, but they have been well identified, and mostly avoided in our implementation.

TOPSIS ideal and anti-ideal concept provided a way to evaluate single scenario cases, and a flexible and easy way to add / remove /change indicators, which is a common situation.

Both methods use simple algebra and therefore can be easily implemented in a spreadsheet.

4.4.1 Methodology caveats

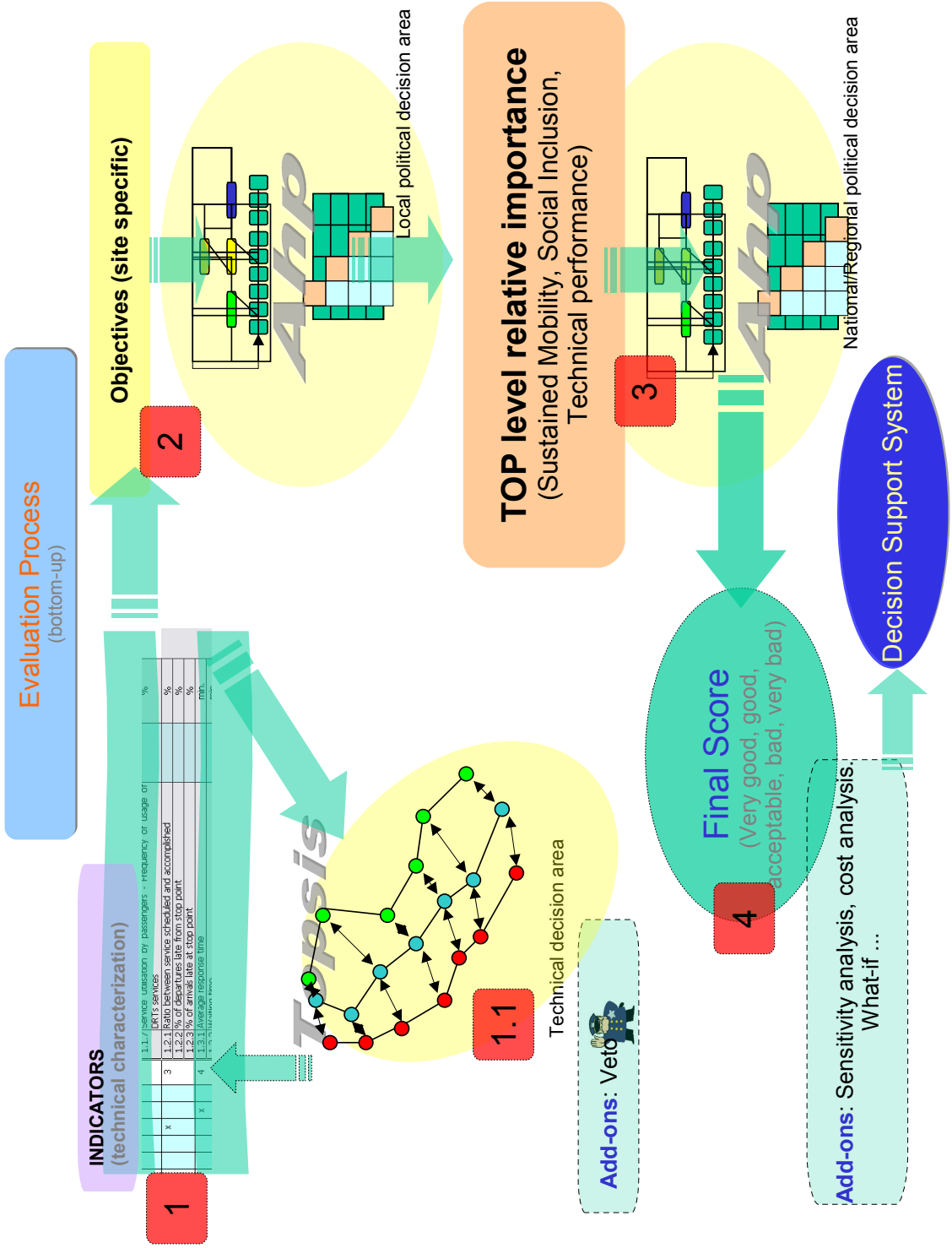
Our proposed framework naturally inherits the problems from the underlying methods. Some were avoided by design, some remain and the concept also creates new problematic areas of its own. Next we list these areas of concern.

1. Users can easily choose to use direct input of weights. This should be avoided as the textual comparison are lost and these are most informative, specially to people not directly involved. Users tend to know their problem very well and therefore might see no immediate advantage in using the AHP method. Some explaining and persuasion might be necessary.
2. All indicators have the same weight. This was not a problem in our particular

implementation, but a situation where this might arrive can be easily foreseen.

3. Numerical indicators are expected to be monotonic. If this should not be the case, the indicator value must be evaluated by other means and its value converted to be used, e.g. in a [1-5] scale.
4. Information is lost in each aggregation. The effect of this loss of information cannot be easily defined. Extra caution is necessary when comparing close values. Also by using a broad evaluation scale as “very bad” to “very good”, attention is needed as some values can be in a frontier situation.
5. The proposed framework was designed with a decision support objective and as a result, all computations can be overridden. This is generally accepted as a “good” characteristic, but, when analysing the results from the evaluation process, one should pay special attention to any overridden computations and look into the reasons for this action.
6. The “ideal” and “anti-ideal” values must be asked to users. The definition of those values must therefore be well understood by them. For example, the “ideal” value (best we can do in some context) might be understood as a “goal” value (value we are aiming at), and this is not desirable.

4.5 Evaluation framework: global view



5 Evaluation Plan

A global evaluation plan was written for the MASCARA consortium. This work has important contributions from all members of the FEUP MASCARA team. The author participated in this work actively, but without any special attribution. This Evaluation Plan explains the methodology and how to use it, and it goes further in defining how to obtain indicators measures. This plan reused material from a previous EU DRT project (SUNRISE). This prior project was successful in defining a pool of indicators applicable to DRT services and a set of questionnaires (and methodology) to help obtaining indicators values. After the consortium agreement on the global evaluation plan, individual evaluation plans where drawn for each site, detailing site specific objectives indicators and questionnaires. This plan summarises the evaluation framework and addresses the practical nature of the necessary work. The last sections of this chapter present a generic guide to the evaluation process. Following this work, the sites then comment on their own evaluation results, producing an "Individual Evaluation Report". Next section presents the fundamental structure and contents of this document. The evaluation methodology is presented addressing its practical implementation.

5.1 MASCARA evaluation approach

This section provides a description of the evaluation methodology designed for the MASCARA project, and its application to the project sites.

DRT services can be designed along a large set of possibilities (both in terms of operational policies as well as transportation means or booking and management procedures) and with a varied number of goals. This means that direct, straightforward comparisons between sites are in general difficult to carry out. Nevertheless this work showed that the developed approach (based on a structured way of organizing and managing information) has a considerable potential in comparatively analysing different cases specially in supporting the assessment and selection of alternatives for the services.

5.1.1 General principles

This evaluation framework was developed for assessing both DRT services already in operation and feasibility projects for DRT services. This methodology explicitly considers the complexity of the processes and the existence of multiple actors and perspectives. For the data collection, some of the work developed in another EU project on DRT – Project SUNRISE (CohéSion sociale dans les zones UrbaiNes/Rurales basée sur services collectifs de mobilité InnovateurS et durablEs)[2] was used, namely a collection of indicators targeting a wide range of DRT

[2] http://srvweb01.softeco.it/sunrise/

functionality, and a set of questionnaires to help gathering this data.

The evaluation is structured along the concepts of high level criteria, objectives, and indicators, reflecting a hierarchical perspective, involving the multiple process stakeholders, and supporting problem structuring and decision-making at different levels, from the more operational to the strategic, political level.

Common to all sites, there are 3 high level criteria – Social Inclusion, Sustainable Mobility, and Technical Performance – which are the pillars of the approach.

Each high level criterion is composed by a set of partial objectives.

Within each objective, a set of indicators were chosen to measure the performance of that particular objective. Sites could design their own indicators or choose from a "pool" of predefined "best practice" indicators, even if with adaptations. This latter option had the advantage of reusing previous work since the methodology and the questionnaires for collecting data were already defined. In particular, indicators from SUNRISE (Deliverable 4 - Common Evaluation Plan: Methodology and Annexes) have been revised and used.

This created a general analysis framework that is basically the same for all the situations, but can be tuned by the selection or addition of specific objectives and indicators.

The approach is supported by an integrated Decision Support System specially developed for this purpose and based on a rather elaborated spreadsheet. The tool (and the evaluation methodology) proved to be flexible enough to be used across DRT scenarios largely different from each other.

As a basic step of this approach, the different sites (both for the demonstration and feasibility studies) have been carefully characterized, their main requirements, expectations and constraints identified, and their general objectives clarified. This work has clearly highlighted an important number of specific features and perspectives in the different sites, but has also shown a large number of common aspects.

In what follows, we will use some summary tables as examples of the main "objectives" chosen by sites for each higher level "criterion". In these tables there are "weights" and "scores" for the objectives.

The weights are assigned by the analysts according to the relative contributions of the objectives towards the fulfilment of each criterion.

Each score is an evaluation on the performance for each criterion or some particular objective. This score is in the [1..5] interval, 1 being a "bad" or poor achievement and 5 an "excellent" performance. The scores ultimately result from measuring the values of indicators (in demonstration sites) or from estimating their values (for feasibility studies).

5.1.2 The evaluation process

The global MASCARA evaluation process for the feasibility and the demonstration

studies is represented in the Figure 20.

The evaluation process involves two main stages:

– DRT Evaluation (a priori) and

– Actual Demo DRT Assessment (the feasibility studies should not consider this stage).

In order to make the evaluation it is necessary to collect some local, regional and national information. It is also necessary to go through a preparation stage, where is expected that each site identifies priority user groups and areas, checks for existing policies and programs and proposes candidates DRT systems to be evaluated.

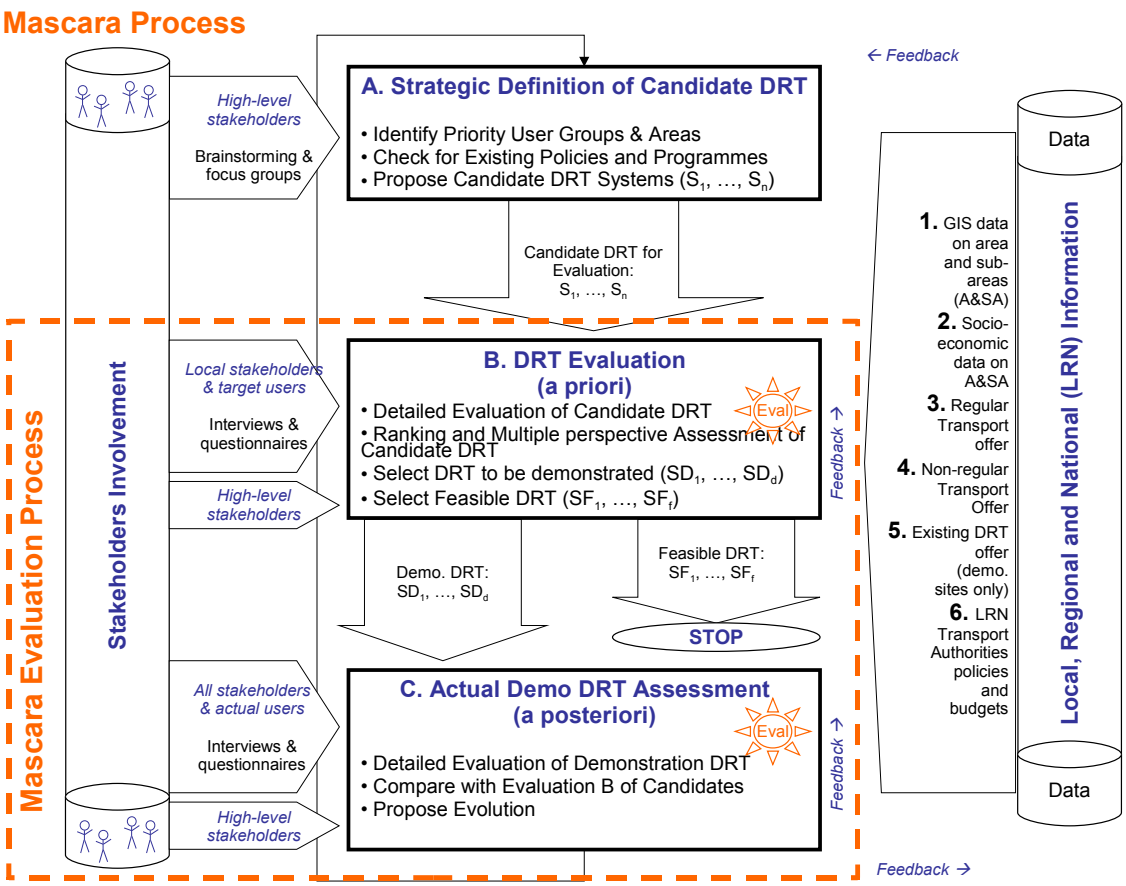


Figure 20: MASCARA Evaluation Process

5.1.3 Global objectives and evaluation model

The main objectives of this project component are:

- to create a common evaluation framework (i.e. a methodology and a set of tools to assess the performance of existing or future DRT systems);
- to provide an assessment of all studies and trials performed in the project, in a

multiple perspective way, taking into account the 3 main, global, high-level criteria.

1	Social Inclusion
2	Sustainable Mobility
3	Technical Performance

Table 3: MASCARA main criteria

The framework should allow:

- an easy (multi-attribute) evaluation and comparison of alternatives;
- an "absolute" assessment of an existing DRT and alternatives;
- the estimation of impacts of a potential solution / policy (simulate "prospective" alternatives for new sites).

The Evaluation Plan (EP) should be viewed as a tool to support investment decisions on DRT systems, or to assess the performance of already existing systems. We assume therefore there is a user and an available procedure to support the decision-making process.

For this purpose, information is used that allows an aggregate, high-level description of the decision environments and of the potentially interesting DRT systems (the effort to collect, to structure and to interpret this information should be "reasonable", and the information should be as aggregate as possible).

For a given environment (site), the required information should result from data produced by the site "actors" (collecting statistical data from various sources), from the answers to specially designed questionnaires (filled by different entities and people) and by expert opinions, probably expressed during some informal interviews.

A set of criteria are proposed (along with attributes that directly or indirectly measure the quality of a given solution for a given site) – these criteria are both qualitative and quantitative (reflecting investment and operation costs, quality of service, environmental impact, sustainability, …) and should allow the user to study and assess any alternative in a multiple perspective way.

The evaluation framework was designed with these objectives in mind. The "evaluation tool" helps to carryout the evaluation work and it was implemented using a spreadsheet. The process can be summarized as a two step approach as described in the following sections.

5.1.4 Set-up (phase 1)

In the "set-up" phase, the following steps are performed:

- for each implementation, the user must first clarify his objectives in the three

main MASCARA criteria, i.e. “Social Inclusion”, “Sustainable Mobility” and “Technical Performance” - this is done by assigning weights to the criteria and a decision support system based on the AHP (Analytical Hierarchical Process) methodology was implemented to help in this process if required;

- within each criteria, (a number of) objectives have to be defined. They are expected to contribute to the associated criteria, and again, weights have to be defined reflecting the relative importance of each objective;
- within each objective, a set of indicators have to be chosen to measure the performance of that particular objective.

In the last set-up step, sites can design their own indicators or chose from a “pool” of predefined “best practice” indicators, even if with adaptations. The latter has the advantage of reusing previous work since questionnaires and methodology for collecting data are already defined. In particular, indicators from a previous EU DRT project (SUNRISE) can be used. This process is illustrated by Figure 21.

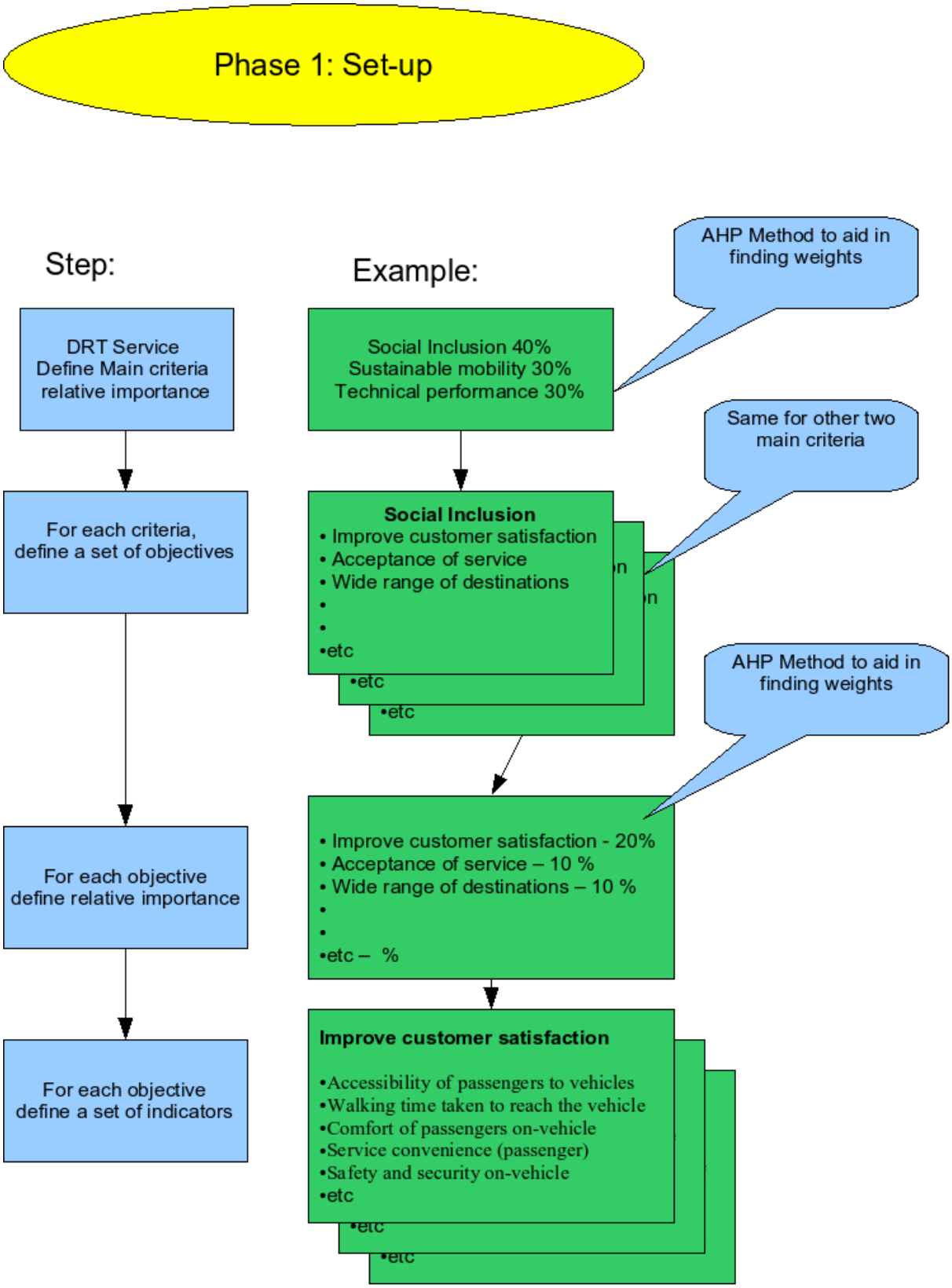


Figure 21: MASCARA Evaluation Process – Phase 1

5.1.5 Data collection and computation (Phase 2)

After the set-up phase, we refer to each objective (to which a set of indicator values has been associated), and a [1..5] score is given by the user, reflecting its performance within those indicator values. Since some indicators can be of a descriptive nature, this score is given as an "expert opinion".

Since most indicators are expected to be of a numerical nature, another decision support methodology (TOPSIS) was implemented for this step. By defining "ideal" and "anti-ideal" values for indicators, an automatic computation can be preformed on the objective score. This is of course intended only to support the actual "expert user" but has two other important benefits: providing additional information to a "non-expert" user, and providing a "check" to the set of indicators. Indeed, if it happens that a computed value and an "expert opinion" differ considerably, then one should question if the set of indicators is actually representing that objective.

After all objectives are scored, a simple computation using the weights given in the set-up phase, yields the scores to each criterion, and even a global score for the DRT solution under analysis can be computed. (in a 1 to 5 scale).

This second phase is illustrated in Figure 22.

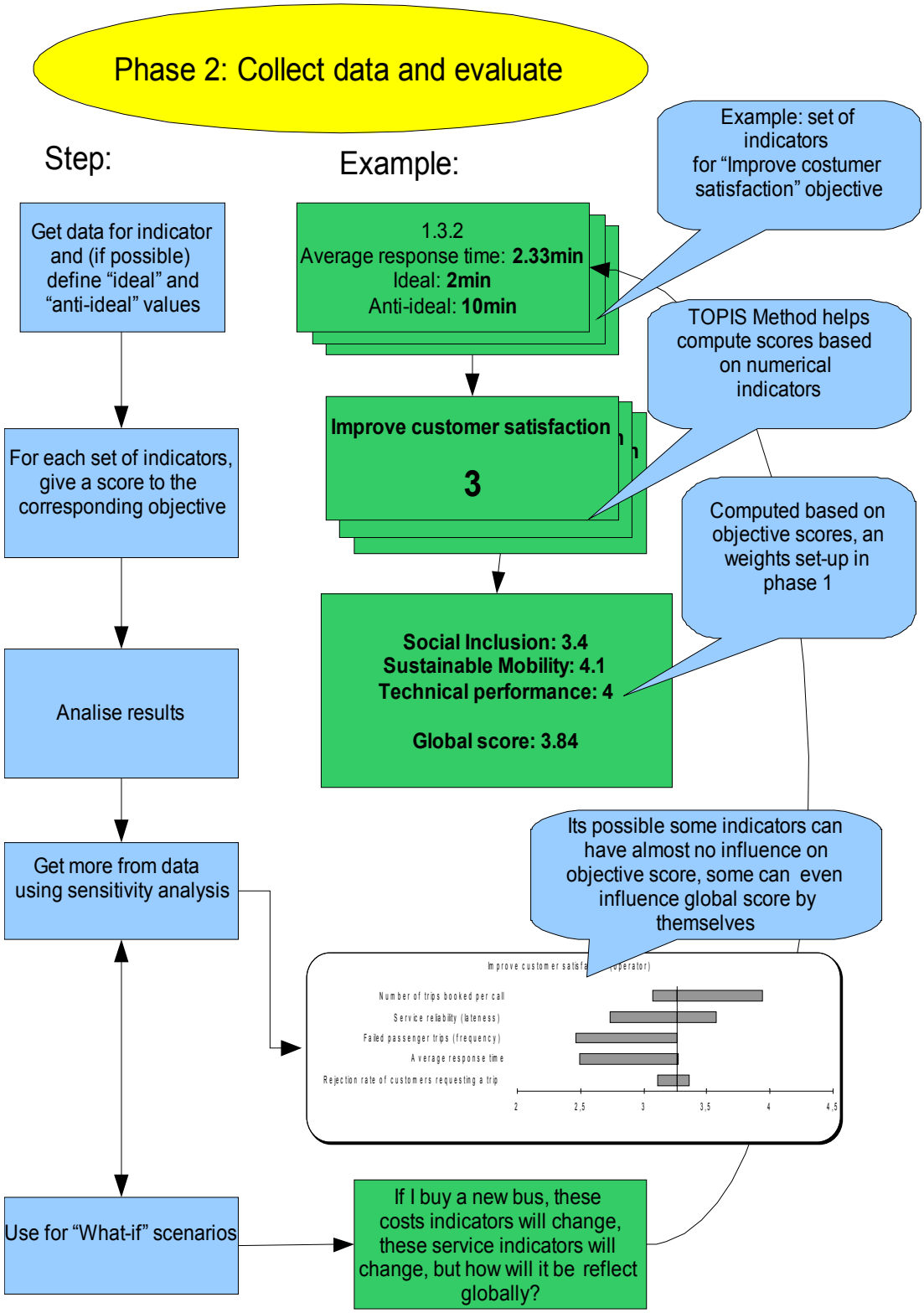


Figure 22: MASCARA Evaluation process - Phase 2

Probably the main value of this semi-automatic approach comes from providing a

way to analyse a complete set of data for a given DRT service. In a feasibility stage, it can be used to compare different scenarios (data variations). Also, after the data is collected it is possible (and somehow trivial) to conduct a sensitivity analysis on the system and also perform different types of "what-if" analysis. After the system has been implemented, the evaluation tool can still be used to evaluate the service performance, and to study or design service changes.

For illustration purposes, figure 23 shows the indicators associated with a given objective, as they appear in the evaluation tool.

OBJECTIVE		Weigh
2.1	**Enlarge Public Transport users**	**20%**

Table 4: Objective and weight example

score	aid		Support Indicator	value	D,M ..	Units
4	2	2.1.5	Load factor	1,21	M	
		2.1.7	Passenger kilometres per vehicle	2650	M	#
		2.1.12c	Number of trips per vehicle trip	1,43	M	#
		2.2.1	Number of bookings per vehicle trip	1,14	M	#
		2.2.2a	Number of calls per day	23	M	#

(figure continues)

,M ..	Units							
M			1,21	2	1	1		
M	#		2650	3000	500	1		
						0		**2,4**
						0		
M	#		1,43	3	1	1		
M	#		1,14	3	1	1		
M	#		23	30	5	1		[0 .. 5]

Figure 23: TOPSIS Calculated objective score based on indicator values

The first column (score) is an (example) expert opinion of the performance of that particular objective.

The second column (aid) is the TOPSIS calculated value

The third column is a descriptive number for the indicator

The fourth column shows the name of that particular indicator

The fifth column is the actual value

The sixth column is an indication of the nature of the indicator. This can be:

D – Descriptive – For textual descriptions when applicable

M – Measured

EL – Estimated locally: estimation based on local knowledge and experience

EO – Estimated based on other information, e.g. other sites experience

G – Set as goal

The following next three columns of this figure are used to show the ideal and anti-ideal values for the associated indicator and if the indicator should be used for the TOPSIS calculation (1) or not (0).

The last column shows the computed, rounded, TOPSIS score based on (numerical) indicator values.

In conclusion, the designed methodology tries to help evaluating DRT services by providing:

1. an assessment tool capable of a service wide evaluation (global, criteria and objective scores) while keeping more detailed information, for detailed analysis (indicator values);
2. a way to evaluate changes to service configurations;
3. a way to continually evaluate a DRT service.

5.2 MASCARA Evaluation procedures and guidelines

This section presents in more detail the process followed by sites. This is not a comprehensive description, since each site followed a specifically designed guidelines, resulting in a document named "Individual Evaluation Plan".

The reader might also notice these procedures propose a different workflow than that presented earlier for the order of phase 1 and phase 2. It is proposed that indicators should be measured first, then the weights for objectives and criteria be found. The reason to this is that the guidelines were written prior to field implementation. It was found when implementing the individual evaluation plans for each site, that it was more practical to address criteria and objectives weights first. The order for the phases are actually interchangeable.

5.2.1 Objectives and general data requirements

We assume there is a local user and an available local procedure to support the decision-making process. For that purpose, information is used, that allows an aggregate, high-level description of the decision environments and of the potentially interesting DRT systems (the effort to collect, to structure and to interpret this information should be "reasonable", and the information should be as aggregate as possible).

For a given environment (site), the required information should result from data produced by the site "actors" (collecting statistical data from various sources), from the answers to specially designed questionnaires (filled by different entities and people) and by expert opinions, probably expressed during some informal interviews.

A set of objectives should be proposed (along with attributes/indicators that directly or indirectly measure the quality of a given solution for a given site) – these objectives are usually of both qualitative and quantitative nature (reflecting investment and operation costs, quality of service, environmental impact, sustainability, …) and should allow the user to assess each in a multiple perspective way.

The Evaluation Process assumes that one or several possible DRT systems have previously been selected (identified) and carefully characterised, according to a set of somehow non-formalised procedures, as described above.

5.2.2 Expected results of the evaluation process

The evaluation process is designed to be used in several ways, which should be adapted to the specific purposes of each situation.

The main generic goal of the process is to assess a given DRT system or concept, in

the case it is being studied for feasibility purposes, or in the case it has already been implemented and is in operation.

These ways are somehow linked to the two different stages at which the analysis can be taken. The first stage is more related to the "technical" characterisation and assessment of solutions. It is therefore more "descriptive" and objective. The second stage involves assigning grades of importance ("weights") to each objective and to each criterion. These judgements should reflect what we might call "politic and strategic objectives" defined by the different stakeholders (operating companies, city administrations; financing bodies, interest groups, …).

Example

Here we consider that a small subset of indicators has been selected by the users (decision-makers), and that these indicators well reflect the key aspects for the decision-making process. The values of these indicators are collected or computed (eventually based on some more detailed information or "sub-indicators).

This (descriptive) model just presents these indicator values in such a way that a global assessment of the concept ("system") is easier.

This might look as follows for criterion 1 (Social Inclusion). A similar small subset of objectives (see following examples) might be computed or collected for the other two high-level criteria, thus creating a "picture" of the system / concept, hopefully easy to assess by the different stakeholders.

1. Social Inclusion			
	Objective and indicators	**definition**	**value**
1.1.	**Service characteristics**		
1.1.4 a	Service reliability (lateness)	Reliability of the service as measured by adherence to estimated departure times given to pre-booked passengers according to: ratio between planned and actual service time; late departure from stop point (% of departures late; average late departure time; reasons); late arrival at stop point (% of arrivals late; average late arrival time; reasons).	*(description)*
1.4.	**Passenger characteristics**		
1.4.1	% of resident population served in area	Percentage of total resident population in service area to which the flexible transport service is available. This is a measure of the extensiveness and coverage of the service.	60
1.4.12	Comfort of passengers at waiting/interchange points	Perception by passengers of comfort at waiting/interchange points (excluding private residences) according to e.g. waiting conditions; staff attitudes.	acceptable
1.4.13	Service convenience (passenger)	Perception by passengers of the convenience of using the service according to e.g. waiting times, accessibility, frequency, route and cost of travel, access to facilities//services/social inclusion.	good

Table 5: Objectives and indicators example for Social Inclusion criterion

2. Sustainable Mobility			
	Objective and indicators	**definition**	**value**
2.1.	**Operational characteristics**		
2.1.7	Passenger kilometres per vehicle	Number of kilometres travelled by passengers per vehicle = total kilometres travelled by all passengers ÷ total number of vehicles. This is a measure of vehicle utilisation.	35
2.1.9	Fare revenue/operating cost	Ratio of fare revenue to cost of vehicle operation.	0.8
2.2.	**TDC**		
2.2.4a	TDC booking and dispatch cost per year	Annual TDC booking and dispatch cost according to: total TDC cost; running cost; maintenance cost; updating cost. *NB: there are many models for organising a TDC, so values for this indicator will be interpreted with care, using the Economic Viability Site Description for the TDC.*	15k€
2.7.	**Political viewpoint**		
2.7.3	Impact on area	The impact of the flexible transport service upon the area according to: car dependence; identification of local transport resources; local business; retaining the local population; non car user tourism; potential development of the area.	*(description)*

Table 6: Objectives and indicators example for Sustainable Mobility criterion

3. Technical Performance			
	Objective and indicators	**definition**	**value**
3.1.	**System speed**		
3.1.7 a	System performance	Mean, maximum and minimum time taken to: search for the customer address; provide conventional public transport information (if applicable)	30s; 75s; 15s
3.3.	**System accuracy / reliability**		
3.3.3	Scheduling reliability (perceived)	Perception by passengers and drivers of compliance of vehicles to the schedule.	good
3.4.	**System acceptance**		
3.4.8a	Driver attitudes towards Travel Dispatch System procedures	Perception by drivers of the TDS procedures according to e.g. start up and switch off procedures; checking and deleting routes; communication with the TDC.	acceptable

Table 7: Objectives and indicators example for Technical Performance criterion

For each indicator, ideal and anti-ideal values should be defined, where the ideal value should be set to the best possible value within the local context, and the anti-ideal value should be set to the worst possible value within the same context. The evaluation tool then presents a computed score based on the actual indicator values and their respective ideal and anti-ideal definitions. This score is an aid to the expert user evaluation.

The user will analyse the computed score regarding the indicator values and decide how good the alternative is for each "objective" (part of a given criterion) indicators. For example, he might say that, given the values of indicators 2.1.7 (Passenger kilometres per vehicle) and 2.1.9 (Fare revenue/operating cost), objective 2.1 (Operational characteristic) achieves a value of "good". This assessment should result from a direct analysis of the user, confirming the TOPSIS computed score. In case the assessment value differs from the computed value, the user should consider and additional explanation.

5.2.3 Relative importance of objectives and criteria

The decision-maker will then focus on the relative importance of the criteria and of the objectives.

Therefore he assigns grades of importance ("weights") to each objective, expressing the "contribution" (relative importance) of each objective to its direct upper level criterion.

This will reflect more strategic concerns and preferences on the concept under analysis.

5.2.4 Scores

The main goal of the process is to assign a value (grade) to each alternative. This value will be a number in the interval [1..5] that can be viewed as a total score or a grade in a 5-levels scale as follows:

1	2	3	4	5
very bad	bad	acceptable	good	very good

Table 8 - Five point score textual equivalence

This grade is obtained simply by transforming the numerical value into a verbal expression, e.g. if the value is 3.5 this means the alternative is "good"; if it is 1.4, it will be considered "bad"

This can be done through indirectly using a pair-wise comparison procedure (of the AHP type) or directly, with a simple statement of the weights for each objective.

5.2.5 The model representation

The model will have assigned weights to objectives and criteria according to the user preferences. After the indicator values are found, a sensitivity analysis will be preformed on the complete data, thus assessing the robustness of the particular alternative under analysis.

In order to incrementally and correctly design the model, a permanent strong feedback from the main site actors is obviously required, along the whole project duration.

To guarantee that the procedure is efficient (and really usable) the users will be required to characterize the geographical zone for which the new DRT system is being studied (or in which an actual system already operates). Moreover information concerning current transportation offer and the demand (i.e. users' requirements) for special transport, should be detailed for the "domain" under consideration (limited geographical area, specific classes of users, habits or economic constraints,…).

5.2.6 Criteria

Three criteria are defined at a rather high level. These are the following:

Social inclusion

Involving "service provision", high-quality flexible transport services, customer satisfaction, opportunities for people to access key destinations, reduction in structural disadvantage for people living in rural areas, frequency of services, increase in effective area for social services and cultural and sports activities.

Sustainable mobility

Including "economic viability", operating costs per passenger, implementation of desired service level within available resources, passengers carried, level of fare-paying.

Technical performance

Capacity of telephone system, advance trip booking time, trip cancellation time, system failures, reliability of data records, accuracy of boarding time, scheduling reliability, frequency of communication, speed of data transmission between on-board and Travel Dispatch Centre (TDC).

The two first criteria do directly reflect MASCARA strategic objectives. The third one tries to capture all specific technological aspects that limit the system performance, and that are not directly represented in the other two.

5.2.7 Objectives

Each "general" criterion is further decomposed into a set of objectives for which a "weight" is assigned, i.e. a level a contribution to the "satisfaction" of the upper level criterion.

1 (Social inclusion)	2 (Sustainable mobility)	3 (Technical performance)
1.1 Service characteristics	2.1 Operational characteristics	3.1 System speed
1.2 Operator characteristics	2.2 Travel dispatch centre	3.2 System failure
1.3 TDC booking characteristics	2.3 Drivers	3.3 System accuracy/ reliability
1.4 Passenger characteristics	2.4 Vehicles	3.4 System acceptance
1.5 Passenger information and marketing	2.5 Statutory providers	3.5 Telecommunications
1.6 Statutory providers	2.6 Political viewpoint	
1.7 Political viewpoint		

Table 9: Main criteria and objectives example

5.2.8 Indicators

Indicators have to be selected for (directly or indirectly) "measuring" the "value" of each aspect important to the fulfilment of each objective. These do not need to be quantitative, but may be expressed in a qualitative (e.g. textual) scale. However for algorithmic purposes, only numerical indicators will be considered during computations.

5.2.9 Process phases and data

The evaluation process considers a proposed solution (possibly from the set of systems under analysis), and works with it separately.

For that given solution, the process is structured around 2 main phases.

Technical phase

A first phase should be "descriptive", objective, and more technical. It is used to characterise the system in a comprehensive way.

Afterwards that characterisation will be used to assign a "value" [1-5] or [very bad, bad, acceptable, good, very good] to each objective. This assessment should be performed by the user (or several expert users), although corroborating (or explaining why not) an auxiliary computation based on indicator values.

Judgement phase

This phase is used to capture the relative importance of criteria and objectives, and involves analysing the grades of importance ("weights") to each objective and to each criterion. This should be based on judgements reflecting the political priorities and preferences of higher level decision-makers. Probably those values should reflect some level of consensus and should result from prior negotiations between the different actors (involved entities and people).

These judgements need to take into account the economical and political values of the decision-making environment. They are “subjective” and “discretionary” by nature. However, the proposed methodology aims at structuring and supporting this process, thus providing decision-makers with a higher level of confidence.

Moreover it is obvious that these judgements should reflect what we might call “strategic objectives” defined by the high-level stakeholders (city administrations; national regulators and financing bodies).

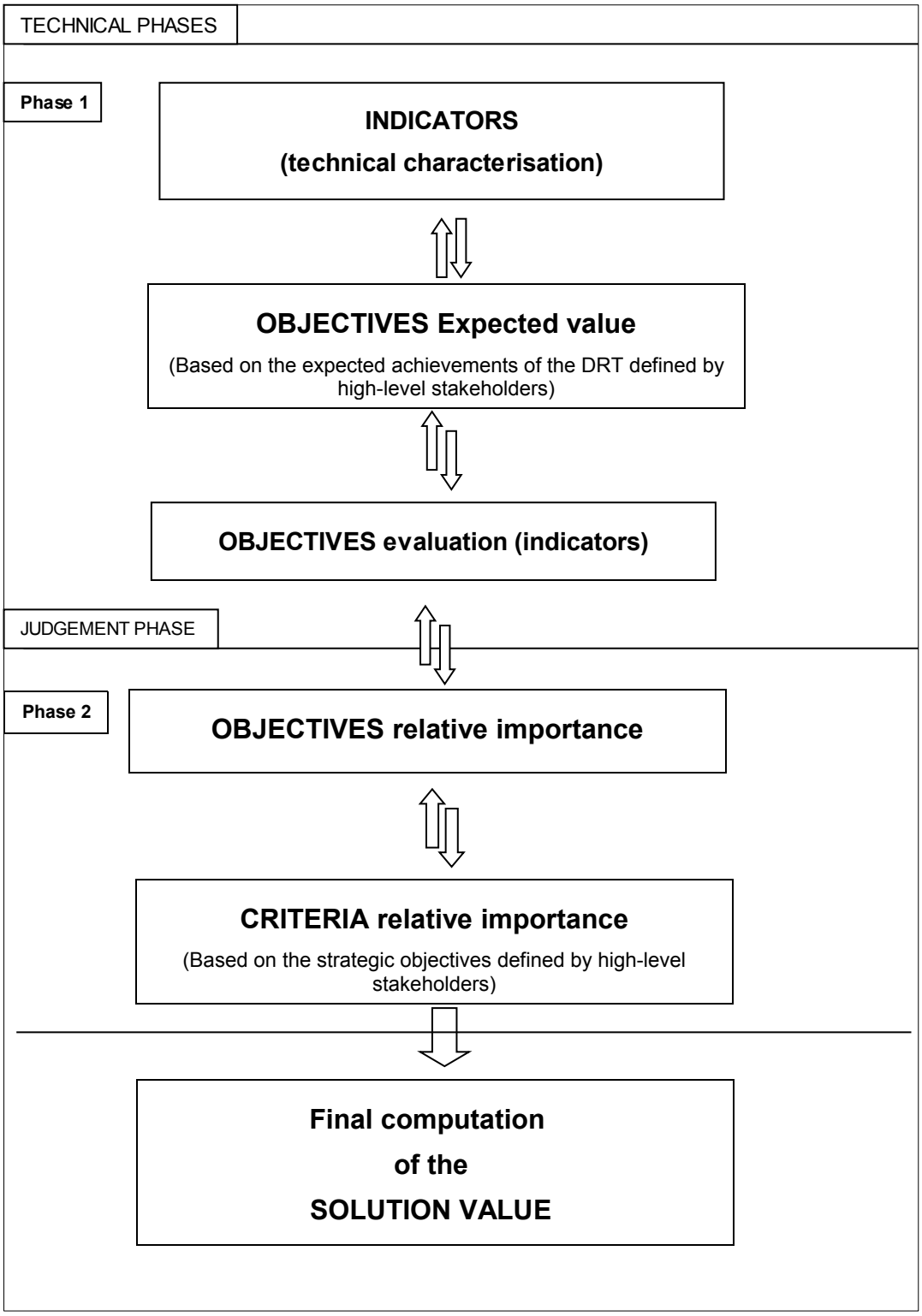


Figure 24: Evaluation Workflow

Phase 1] Technical characterisation of the DRT

A set of support indicators has to be defined. Some of these indicators are quantitative and some are qualitative (they are expressed in different scales and measures).

Some indicators are not relevant for some cases – then they should be ignored.

For identification purposes, indicators are numbered (X.Y.Z), e.g. indicator 3.1.5 is "Trip handling time".

Indicators are related to the evaluation criteria (X) and objective (Y). The above example indicator (3.1.1) is related to global criterion 3 ("Technical Performance") and to objective 3.1 ("System speed").

A predefined set of indicators might be used (reflecting the inputs from the partners and some results of previous projects). Some of these indicators could still need to be made more precise (probably needing to be subdivided).

Sites should develop their own specific indicators (or adapt existing ones) as needed to better describe their own situation.

Objectives evaluation

- Indicators are partitioned into small groups that are related (or associated) to each objective. For example, indicators 3.1.1, 3.1.2, 3.1.3, 3.1.4, 3.1.5, and 3.1.6, are associated with objective 3.1 ("System speed"). For each indicator,ideal and anti-ideal values must be defined.
- For each objective, a score is computed based on the values of the numerical indicators. The USER must analyse the indicator values and decide how good the solution is for that objective, confirming or correcting the computed score. This will be done separately for each objective. For example, he might say that the solution is "acceptable" for objective "Dispatching centre performance" (or equivalently, grade it grade 3).
- The user will enter a short analysis on the value. This description gains importance if the user chooses not to accept the computed value.

Phase 2] criteria and objective relative importances

- Assign grades of importance ("weights") to each objective, expressing the "contribution" (relative importance) of each objective to its direct upper level criterion.
- Assign "weights" to each criterion, based on judgements reflecting more strategic concerns and preferences.
- This can be done through a direct simple statement of the weights, or indirectly using a pair-wise comparison procedure (of the AHP type) – see details below.

End] Getting the final value of the DRT evaluation

Finally, a "double" weighted average of all indicator values (grades), as obtained in phase 1, is computed, thus yielding the final value of the DRT system under analysis.

5.3 Data to be collected by each site

We now present in more detail the requested information to each site. These requests are presented in a formatted and regulated form and constitute a starting point to the evaluation process.

5.3.1 Site Description

Information about each site is required, under the following headings:

- Key geographical characteristics
- Core socio-economic information
- Mobility and infrastructure characteristics
- Description of statutory authorities (including transport, education, social services and health departments)
- Description of operators (including commercial, voluntary/community operators and taxi operators)
- Description of current public transport services
    - Rail
    - Bus services with no user restrictions
    - Bus services with user restrictions (e.g. education, social services, patient transport)
    - Bus services operated by community or voluntary groups
    - Taxis

ACTION

A list of information that should be gathered by each site should be produced. Each partner then proceeds to gather this information and eventually adding to the list when necessary.

5.3.2 Identification of Stakeholders

This information is used to identify User Groups (stakeholders), i.e. those organisations and individuals who have a direct interest in the outcome of introducing the flexible transport service.

The main categories are:

0. Public transport providers
   - 0.1. Statutory authorities (including transport, education, social services and health departments)
   - 0.2. DRT service operator
   - 0.3. Other bus operators (including commercial, voluntary/community operators)
   - 0.4. Rail operators
   - 0.5. Taxi operators
1. Government (elected representatives)
   - 1.1. Local authorities
   - 1.2. Regional authorities
   - 1.3. National government
   - 1.4. EC
2. End users
3. Community representatives
4. Travel Dispatch Centre
5. Businesses
   - 5.1. Software suppliers
   - 5.2. Hardware suppliers
   - 5.3. SMEs

Not all these categories need be represented at each site, e.g. there may be no community representatives with an interest in the site.

ACTION

Each site should identify its own stakeholders. To this task some generic questionnaires can be used as an aid.

5.3.3 Identification of Expected Impacts and Indicators for Site Objectives

Site Objectives

The Site Objectives describe the general aims and operational objectives of the local project, e.g. reduction of private traffic and pollution, rationalisation of the resources for mobility, improving the quality of life in specified areas, increasing social

cohesion, developing services for disadvantaged categories such as the disabled and elderly.

The Site Objectives will be based upon those firstly identified by each site, but may be extended or altered as a result of the Feasibility Studies.

The identification of expected impacts is a key part of the evaluation process as it provides a reference point against which the impacts of introducing a DRT service can be assessed.

The following method is applied:

- For each site objective that has been identified a weight is attributed. This weight can be found by use of the AHP method which promotes a pair wise comparison of the relative importance of the objectives. The user can also enter these weights directly. Weights reflect the importance attached to achieving this objective on a scale of 1 (low) to 5 (high).
- For each objective the expected classification value is given. This can be negative (1 – Very bad or 2 - Bad), no noticeable effect (3 - Acceptable), or positive (4 – Good, Very Good). It is possible that some aspects of the service will have a negative classification, e.g. some aspects of passenger satisfaction may be traded off as part of the flexible transport scheme. The expected value does not necessarily directly relate to the criteria weight level, e.g. the objective may have a low weight but could still have a high positive classification value.
- The expected classification value of the objective is justified by briefly describing the reason(s) for the expectations.
- For each objective, the relevant indicators to be measured are identified.

ACTION

Each site should complete a survey supplied in order to define its objectives and expected impacts. This survey also requires the sites to select a list of indicators that should be measured in this Evaluation Process. The indicators are used to measure site performance according to three Criteria.

Within the following Criteria the indicators are grouped as follows:

Criteria	Indicators group
1. Social inclusion	• Service characteristics
	• Operator characteristics
	• TDC booking characteristics
	• Passenger characteristics
	• Passenger information and marketing
	• Statutory providers
	• Political viewpoint
2. Sustainable mobility	o Operational characteristics
	o Travel dispatch centre
	o Drives
	o Vehicles
	o Statutory providers
	o Political viewpoint
3. Technical performance	1.1 System speed
	1.2 System failure
	1.3 System accuracy/ reliability
	1.4 System acceptance
	1.5 Telecommunications

Table 10: Objectives and indicators example

5.3.4 Indicator pool

A generous number of indicators were defined in a prior EU project (SUNRISE). These indicators can be adapted and used for measurement at the sites, in order to cater for individual site requirements and capabilities.

Sites can also propose additional indicators that would contribute to the evaluation, e.g. the existing list of indicators may not be relevant to one or more of the User Groups that have been identified such as local businesses.

The precise wording of the indicators should be adjusted to suit local site requirements.

A comprehensive list of indicators should be considered by each site. Then they should go carefully trough this list and choose what indicators are more suitable to its case.

5.3.5 Individual Evaluation Plans

The Individual Evaluation Plans (IEP) should be built according to the information gathered in the previous phases of the Evaluation Process.

The IEP should consider each of the following features:

- Strategic objectives
- Target area (primary and secondary)
- Stakeholders (primary and secondary)
- Problems to be solved in the area (mobility needs are not met by the present public transport services, for example)
- Positive points in the chosen area
- Outlined DRT concept
- Tactical objectives for DRT and for the target area
- Evaluation objectives
- Expected outcomes of a DRT system for the target area
- Primary evaluation needs

5.3.6 Evaluation process results

Site Performance

The following information should be provided about each indicator to guide the data collection:

- The reference number of each indicator.
- The source from which the data will be collected. Data for some indicators ….
  - Must be collected from 1 source, e.g. Labour Productivity requires data from the Travel Dispatch Centre.
  - Can be collected from a choice of sources (which ever is the most appropriate), e.g. Formal Training Cost requires data from either the operator or the statutory authority.
  - Must be collected from a combination of 1 or more sources, e.g. Operating Cost per Revenue Hour requires data partly from the operator, on-board the vehicle and the Travel Dispatch Centre.
  - Must be collected from more than 1 source, e.g. Failed Passenger Trips (frequency) requires data from all Travel Dispatch Centre staff and drivers.
- The sample size and duration of data collection (if relevant).
- The components of the indicator including the units of measurement. Some indicators may be collected in one of two ways, according to the availability of data.
- A definition of the indicator.
- An indication if any components of the indicator are required by other indicators.
- Supplementary data collection guidance notes.

Once the indicators have been selected, the appropriate survey forms are chosen and collated according to the following sources:

- Statutory authorities
- Operators
- Drivers
- Travel Dispatch Centre staff
- Passengers
- Politicians
- Travel Dispatch Centre monitoring
- On-board vehicle monitoring
- Others that may be identified by individual sites

Where data are required from more than one source, the survey forms are repetitive.

Once the field survey data have been collected it is reported for each Criterion in Site Performance tables for the indicators that have been selected are constructed. The objective evaluation results are also reported here.

Individual evaluation report

This results and analysis should be presented in a separate document, namely an "Individual Evaluation Report" written by the local stakeholders, where all necessary local analysis on the results should be presented.

6 Evaluation Tool example use

This chapter illustrates the use of the evaluation tool. The tool was constructed using a spreadsheet. Although this spreadsheet uses some advanced features and calculations, it does not include any programming script or macros, therefore it can be used in any kind of spreadsheet software.

Using a spreadsheet software is not optimal, but highly effective since all the stakeholders have some knowledge on the use of spreadsheets. We chose not to use "advanced" features like scripts or macros. They are not necessary for the calculations, and in this way the tool could be used in a wider range of spreadsheet software. Although we chose not to use scripting, some advanced formatting techniques were used to help with user input.

6.1 Guided screenshots

The "book" included a special "Read Me" sheet with general instructions on how to use the tool. This is (partly) shown in Figure 25.

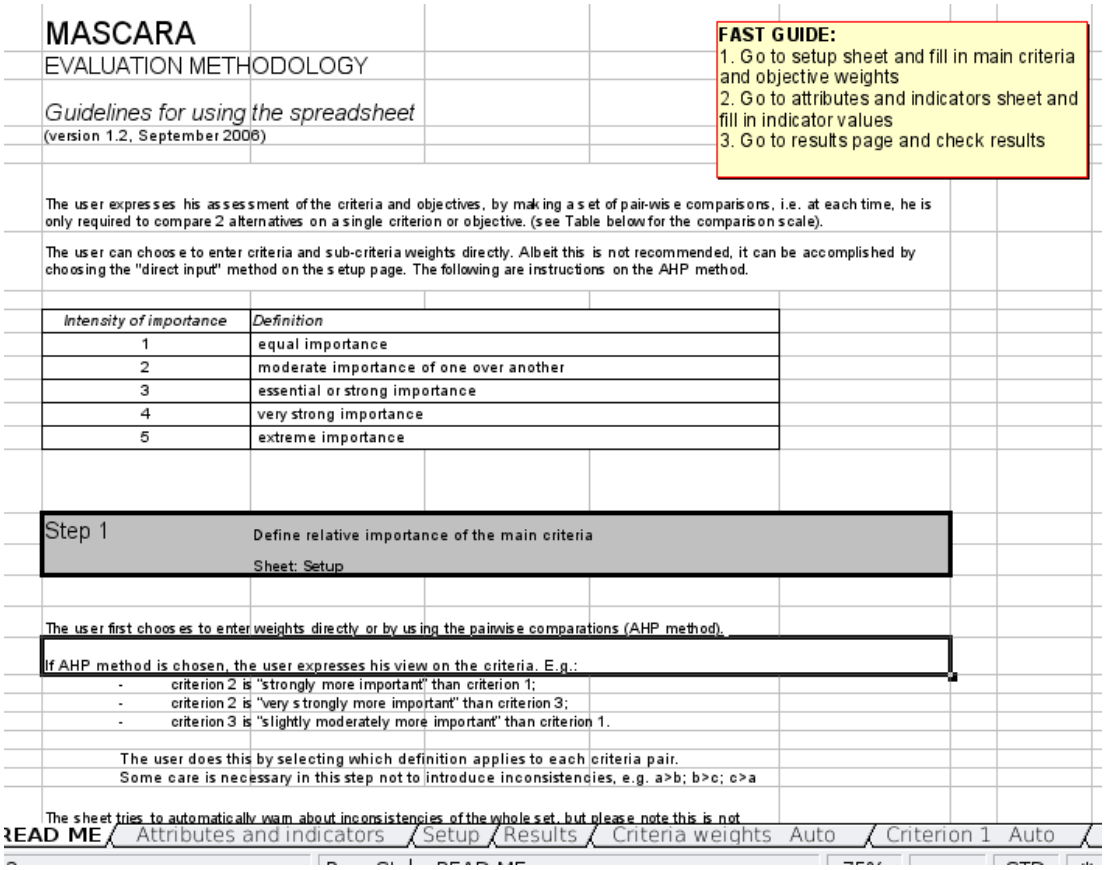


Figure 25: Evaluation tool. Instructions

The first use page, or work entrance point is the “Set-Up” sheet where weights for criteria and objectives should be entered. In this sheet the user is first presented with a choice either to use the AHP methodology or to enter the weights directly. The number of objectives had to be set-up before hand, but the user can change their names.

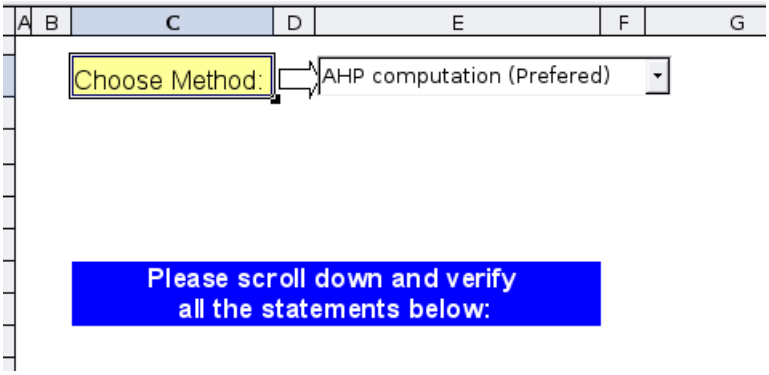


Figure 26: Choosing how to assign weights

If the user chose to “enter the weights directly”, then the screen would change (figure 27).

Direct Input of Critiria Weights

Please Enter criteria and

Editable cells in color

Criteria	Weight %
SOCIAL INCLUSION	60%
SUSTAINABLE MOBILITY	30%
TECHNICAL PERFORMANCE	10%
CONTROL SUM:	100% OK

Criterion 1 SOCIAL INCLUSION	Weight %
Relevant and attractive service	18%
Acceptence of service	18%
Wide range of destinations	8%
Reduce walking time for people with reduced mobility	8%
Service according to quality and performance parameters	8%
Visible increase in societal participation for target groups	15%
Well-distributed user profile	10%
Offer alternatives to private car use	15%
CONTROL SUM:	100% OK

sub-criteria weights below

Criteria 2 SUSTAINABLE MOBILITY	Weight %
Meet minimum funding requirements	10%
Future cost recovery=50%	10%
Cost of booked trip <30% overall costs or 1€	10%
Patronage of services	15%
50% fare paying custumers	10%
Avoid reduction of passengers on the existing Bus Eireann routes in the target area	10%
Establish a stable institutional and organisational framework for the DRTservices	10%
Enable a decision within 2 years by Cork City about DRT future	25%
CONTROL SUM:	100% OK

Criteria 3 TECHNICAL PERFORMANCE	Weight %
Financial and practical solution for TDC centre	25%
TDC offers clear and timely solution to requests	15%
Optimise bookings / dispatches	25%
Optimize number of booked/realised trips	35%
CONTROL SUM:	100% OK

Figure 27: Direct assignment of weights

As previously mentioned, the direct assignment of weights is possible but not desirable, and users are encouraged to use the AHP methodology. In this case, the users are directed to complete a set of comparisons to describe their particular situation. An example is shown in Figure 28.

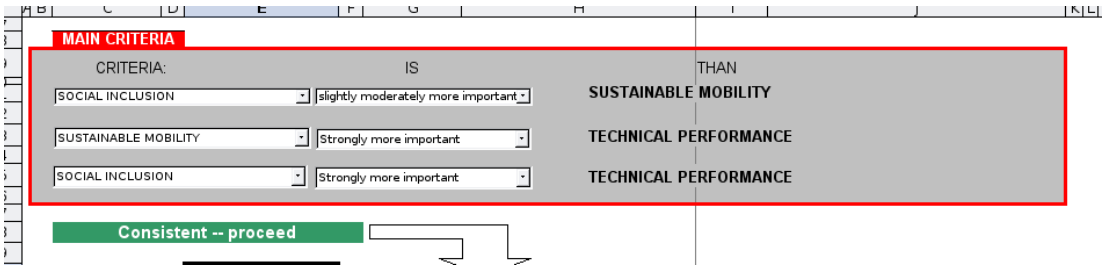


Figure 28: AHP comparisons

A consistency check is immediately presented to the user. The AHP methodology allows for some amount of inconsistency in user decisions figure 29 and figure 30 illustrate this, using the main criteria.

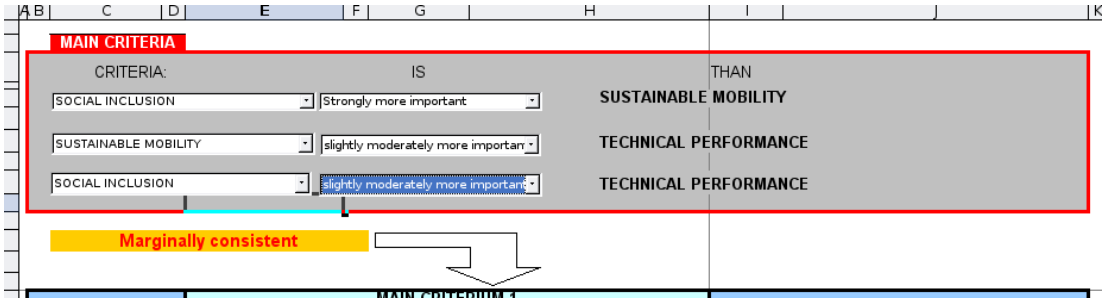


Figure 29: Consistency check example: Marginally consistent

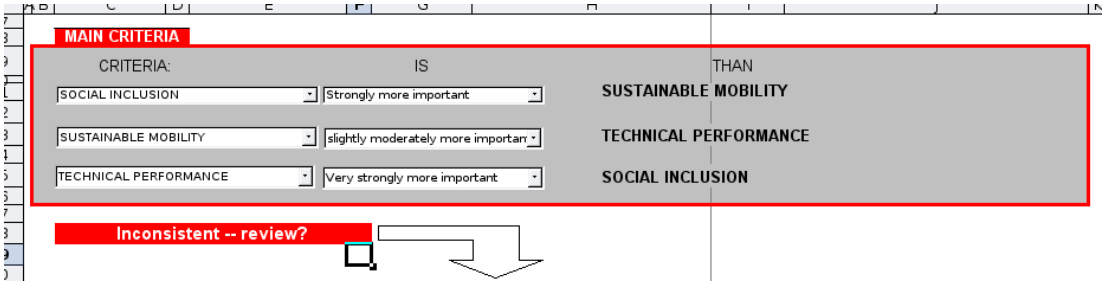


Figure 30: Consistency check example: Inconsistent

Depending on the number of objectives, a table is constructed where all the necessary pairwise comparisons to the AHP method are presented to the user (figures 31 and 32).

MAIN CRITERIUM 1
SOCIAL INCLUSION

OBJECTIVE	IS	THAN
Acceptence of service	Equaly important	Relevant and attractive service
Relevant and attractive service	Strongly more important	Wide range of destinations
Relevant and attractive service	Equaly important	Reduce walking time for people with reduced mobility
Relevant and attractive service	Equaly important	Service according to quality and performance parameters
Relevant and attractive service	slightly moderately more important	Visible inccrease in societal participation for target groups
Relevant and attractive service	slightly moderately more important	Well-distributed user profile
Relevant and attractive service	Extremely more important	Offer alternatives to private car use
Acceptence of service	slightly moderately more important	Wide range of destinations
Acceptence of service	slightly moderately more important	Reduce walking time for people with reduced mobility
Acceptence of service	Strongly more important	Service according to quality and performance parameters
Acceptence of service	Very strongly more important	Visible inccrease in societal participation for target groups
Acceptence of service	Very strongly more important	Offer alternatives to private car use
Reduce walking time for people with reduc	slightly moderately more important	Wide range of destinations
Service according to quality and performar	Strongly more important	Wide range of destinations
Wide range of destinations	Equaly important	Visible inccrease in societal participation for target groups
Well-distributed user profile	Equaly important	Wide range of destinations
Wide range of destinations	Strongly more important	Offer alternatives to private car use
Service according to quality and performar	Extremely more important	Reduce walking time for people with reduced mobility
Visible inccrease in societal participation for	Equaly important	Reduce walking time for people with reduced mobility
Reduce walking time for people with reduc	Strongly more important	Well-distributed user profile
Reduce walking time for people with reduc	Extremely more important	Offer alternatives to private car use
Service according to quality and performar	slightly moderately more important	Visible inccrease in societal participation for target groups
Service according to quality and performar	Strongly more important	Well-distributed user profile
Service according to quality and performar	Strongly more important	Offer alternatives to private car use
Visible inccrease in societal participation for	slightly moderately more important	Well-distributed user profile
Visible inccrease in societal participation for	Very strongly more important	Offer alternatives to private car use
Well-distributed user profile	Strongly more important	Offer alternatives to private car use

Consistent -- proceed

Figure 31: Pairwise comparisons for Social Inclusion criterion (example)

The two remaining criteria are handled in the same manner.

MAIN CRITERIUM 2
SUSTAINABLE MOBILITY

OBJECTIVE	IS	THAN
Meet minimum funding requirements	Very strongly more important	Future cost recovery=50%
Meet minimum funding requirements	Strongly more important	Cost of booked trip <30% overall costs or 1€
Patronage of services	Strongly more important	Meet minimum funding requirements
Meet minimum funding requirements	Very strongly more important	50% fare paying custumers
Meet minimum funding requirements	Strongly more important	Avoid reduction of passengers on the existing Bus Eireann routes in the target area
Establish a stable institutional and organis	Strongly more important	Meet minimum funding requirements
Meet minimum funding requirements	slightly moderately more important	Enable a decision within 2 years by Cork City about DRT future
Cost of booked trip <30% overall costs or	slightly moderately more important	Future cost recovery=50%
Patronage of services	slightly moderately more important	Future cost recovery=50%
50% fare paying custumers	Equaly important	Future cost recovery=50%
Future cost recovery=50%	slightly moderately more important	Avoid reduction of passengers on the existing Bus Eireann routes in the target area
Future cost recovery=50%	Equaly important	Establish a stable institutional and organisational framework for the DRT services
Enable a decision within 2 years by Cork C	Strongly more important	Future cost recovery=50%
Patronage of services	slightly moderately more important	Cost of booked trip <30% overall costs or 1€
Cost of booked trip <30% overall costs or	slightly moderately more important	50% fare paying custumers
Cost of booked trip <30% overall costs or	Equaly important	Avoid reduction of passengers on the existing Bus Eireann routes in the target area
Cost of booked trip <30% overall costs or	Strongly more important	Establish a stable institutional and organisational framework for the DRT services
Cost of booked trip <30% overall costs or	slightly moderately more important	Enable a decision within 2 years by Cork City about DRT future
Patronage of services	Strongly more important	50% fare paying custumers
Patronage of services	slightly moderately more important	Avoid reduction of passengers on the existing Bus Eireann routes in the target area
Patronage of services	slightly moderately more important	Establish a stable institutional and organisational framework for the DRT services
Patronage of services	Strongly more important	Enable a decision within 2 years by Cork City about DRT future
50% fare paying custumers	slightly moderately more important	Avoid reduction of passengers on the existing Bus Eireann routes in the target area
50% fare paying custumers	slightly moderately more important	Establish a stable institutional and organisational framework for the DRT services
50% fare paying custumers	slightly moderately more important	Enable a decision within 2 years by Cork City about DRT future
Avoid reduction of passengers on the exist	slightly moderately more important	Establish a stable institutional and organisational framework for the DRT services
Enable a decision within 2 years by Cork C	slightly moderately more important	Avoid reduction of passengers on the existing Bus Eireann routes in the target area
Establish a stable institutional and organisa	slightly moderately more important	Enable a decision within 2 years by Cork City about DRT future

Marginally consistent

CRITERIA 3
TECHNICAL PERFORMANCE

OBJETIVE	IS	THAN
Financial and practical solution for TDC cer	Equaly important	TDC offers clear and timely solution to requests
Financial and practical solution for TDC cer	Equaly important	Optimise bookings / dispatches
Financial and practical solution for TDC cer	slightly moderately more important	Optimize number of booked/realised trips
Optimise bookings / dispatches	Equaly important	TDC offers clear and timely solution to requests
TDC offers clear and timely solution to req	Strongly more important	Optimize number of booked/realised trips
Optimise bookings / dispatches	slightly moderately more important	Optimize number of booked/realised trips

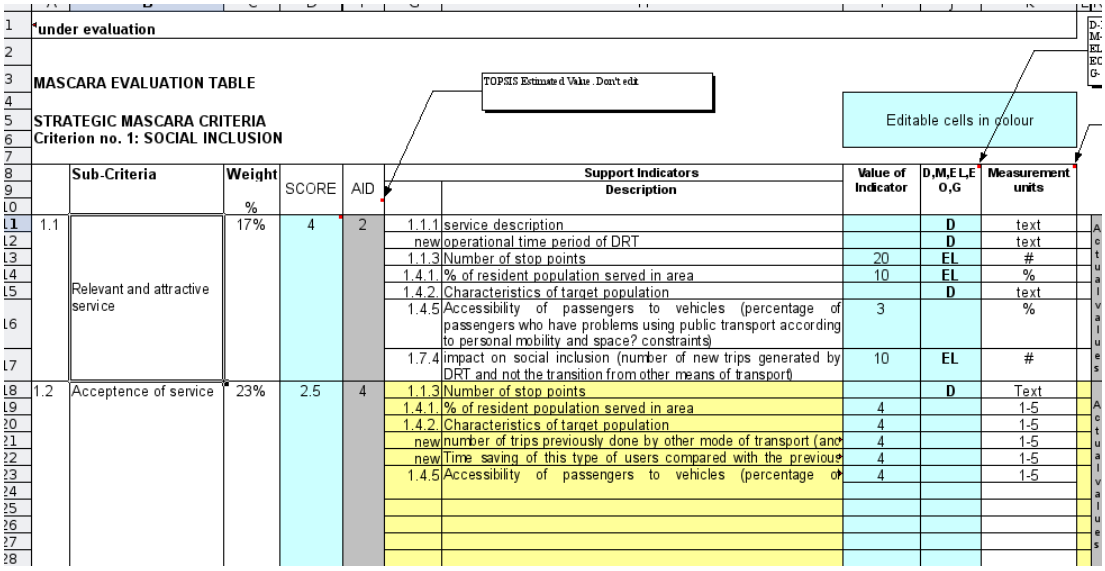
Consistent -- proceed

PLEASE REVIEW DECISIONS!!! GO TO RESULTS ONLY IF SURE

Figure 32: Sustainable mobility and Technical performance criteria (example)

At this stage, indicator values could have been already collected. If this was the case, the user would start looking at the results at this stage. In MASCARA project, all of the sites did this step of the evaluation first.

In this case, the next step is to collect and enter information about the indicator values, in the "Attributes and indicators" page. Figures 33 and 34 illustrate this process. This spreadsheet page is expected to contain a large amount of information. For illustration purposes, we just show an excerpt.



under evaluation

MASCARA EVALUATION TABLE

STRATEGIC MASCARA CRITERIA
Criterion no. 1: SOCIAL INCLUSION

TOPSIS Estimated Value . Don't edit

Editable cells in colour

	Sub-Criteria	Weight %	SCORE	AID	Support Indicators	Description	Value of Indicator	D,M,E L,E O,G	Measurement units
1.1	Relevant and attractive service	17%	4	2	1.1.1	service description		D	text
					new	operational time period of DRT		D	text
					1.1.3	Number of stop points	20	EL	#
					1.4.1.	% of resident population served in area	10	EL	%
					1.4.2.	Characteristics of target population		D	text
					1.4.5	Accessibility of passengers to vehicles (percentage of passengers who have problems using public transport according to personal mobility and space? constraints)	3		%
					1.7.4	impact on social inclusion (number of new trips generated by DRT and not the transition from other means of transport)	10	EL	#
1.2	Acceptence of service	23%	2.5	4	1.1.3	Number of stop points		D	Text
					1.4.1.	% of resident population served in area	4		1-5
					1.4.2.	Characteristics of target population	4		1-5
					new	number of trips previously done by other mode of transport (anc	4		1-5
					new	Time saving of this type of users compared with the previous	4		1-5
					1.4.5	Accessibility of passengers to vehicles (percentage of	4		1-5

Figure 33: Indicators data sheet (picture continues)

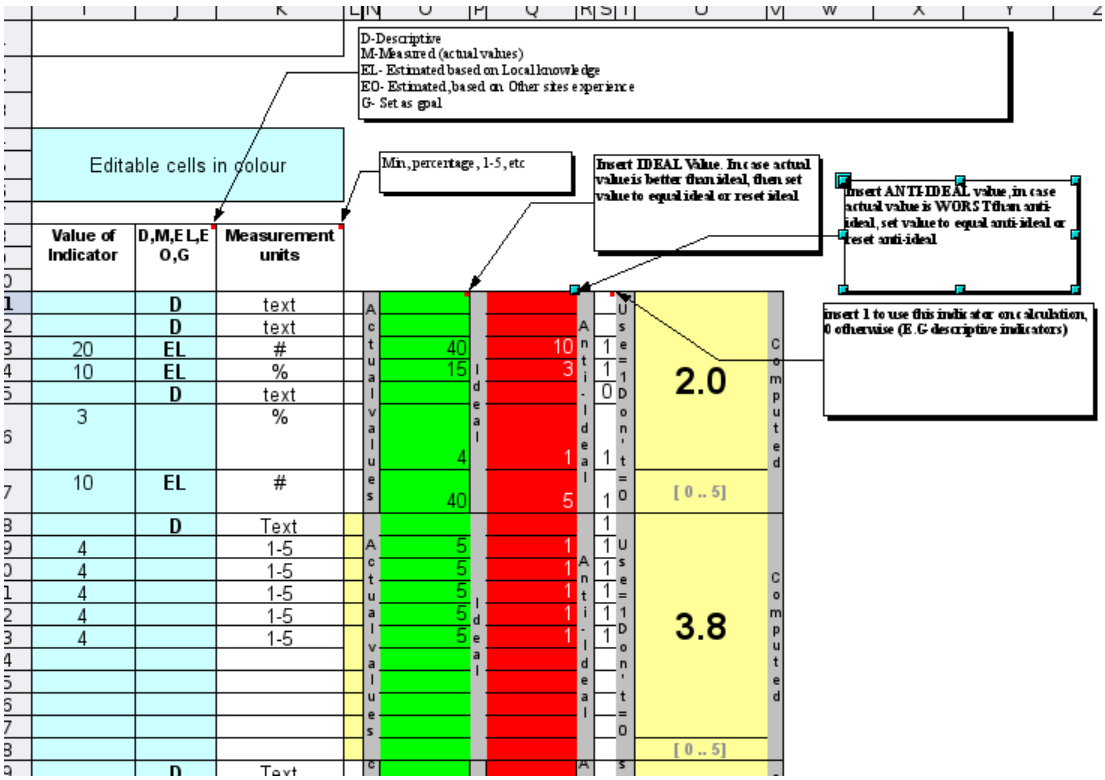
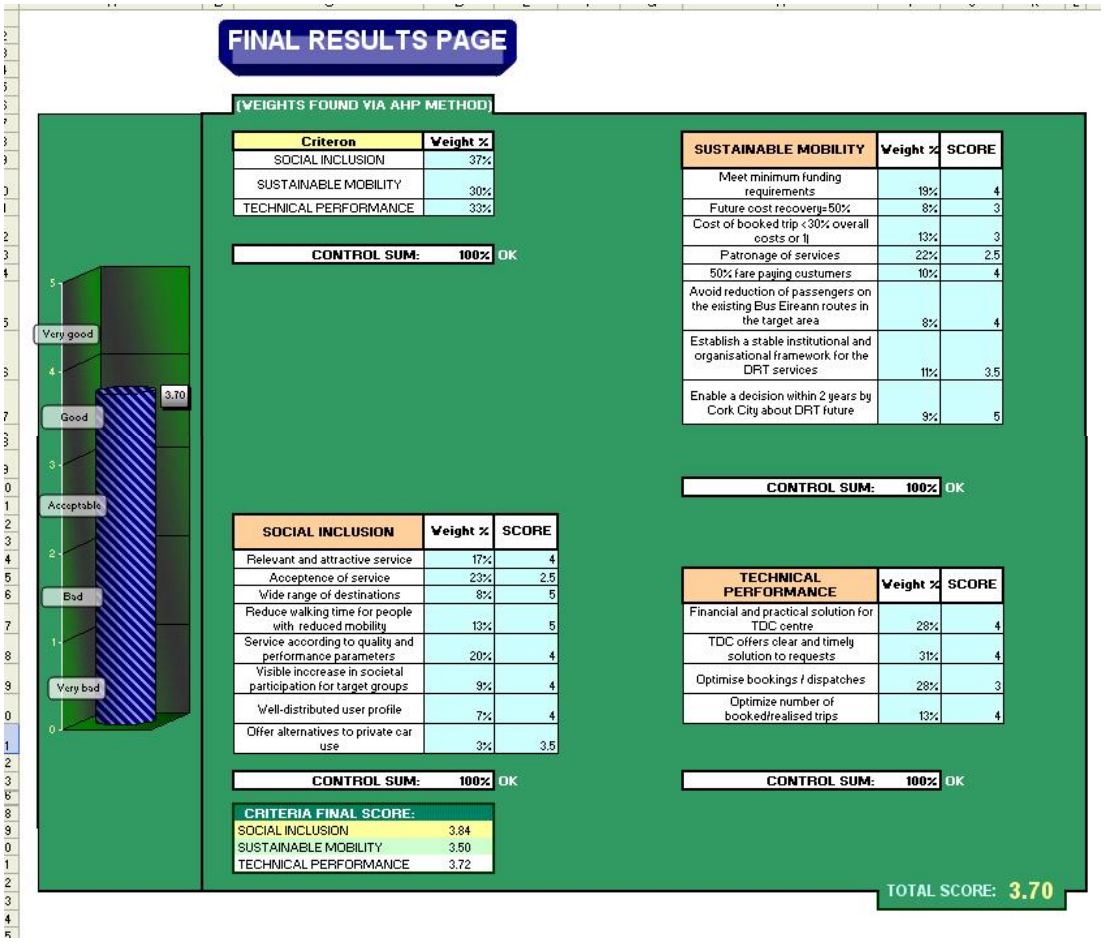


Figure 34: Indicator values data and TOPSIS Evaluation

After the collection indicator values and the assignment of objective and criteria weights, the results are presented in the "Results sheet", shown in figure 35.



FINAL RESULTS PAGE

(WEIGHTS FOUND VIA AHP METHOD)

Criteron	Weight %
SOCIAL INCLUSION	37%
SUSTAINABLE MOBILITY	30%
TECHNICAL PERFORMANCE	33%
CONTROL SUM:	100% OK

SUSTAINABLE MOBILITY	Weight %	SCORE
Meet minimum funding requirements	19%	4
Future cost recovery=50%	8%	3
Cost of booked trip <30% overall costs or 1)	13%	3
Patronage of services	22%	2.5
50% fare paying custumers	10%	4
Avoid reduction of passengers on the existing Bus Eireann routes in the target area	8%	4
Establish a stable institutional and organisational framework for the DRT services	11%	3.5
Enable a decision within 2 years by Cork City about DRT future	9%	5
CONTROL SUM:	100%	OK

SOCIAL INCLUSION	Weight %	SCORE
Relevant and attractive service	17%	4
Acceptence of service	23%	2.5
Wide range of destinations	8%	5
Reduce walking time for people with reduced mobility	13%	5
Service according to quality and performance parameters	20%	4
Visible increase in societal participation for target groups	9%	4
Well-distributed user profile	7%	4
Offer alternatives to private car use	3%	3.5
CONTROL SUM:	100%	OK

TECHNICAL PERFORMANCE	Weight %	SCORE
Financial and practical solution for TDC centre	28%	4
TDC offers clear and timely solution to requests	31%	4
Optimise bookings / dispatches	28%	3
Optimize number of booked/realised trips	13%	4
CONTROL SUM:	100%	OK

CRITERIA FINAL SCORE:	
SOCIAL INCLUSION	3.84
SUSTAINABLE MOBILITY	3.50
TECHNICAL PERFORMANCE	3.72

TOTAL SCORE: 3.70

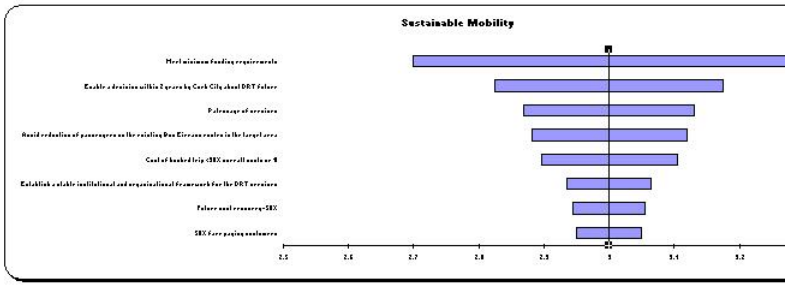
Figure 35: Evaluation tool - Results page (example)

After all data has been entered, this spreadsheet page shows the aggregated results. All computations take place almost instantly, so, at this stage, the tool can be used to evaluate "What-if scenarios" by changing some data and looking at the results.

Some sensitivity analysis is also available to the user by means of Tornado plots (see section 4.2.3 on page 32).

Some Tornado plots are already defined, but the user is free to generate any other analysis. This is possible since no scripting was used in any calculations, so any spreadsheet "add-in" such as for Tornado plots, are readily usable.

10	Parameter	Output D4(	Output D4(	Output Differences:
11	Meet minir	2.698298	3.301702	0.603404
12	Enable a de	2.824693	3.175307	0.350614
13	Patronage	2.869721	3.130279	0.260558
14	Avoid redu	2.881044	3.118956	0.237913
15	Cost of bo	2.895608	3.104392	0.208784
16	Establish a	2.936161	3.063839	0.127678
17	Future cos	2.944899	3.055101	0.110202
18	50% fare pa	2.949576	3.050424	0.100849
19				
20				
21				
22				
23	Parameter	Output D3	Output D3	Output Differences:
24	Relevant a	2.737491	3.262509	0.525018
25	Acceptenc	2.737491	3.262509	0.525018
26	Service ac	2.864158	3.135842	0.271684
27	Reduce wa	2.886217	3.113783	0.227566
28	Visible inc	2.943131	3.056869	0.113738
29	Well-distrib	2.943131	3.056869	0.113738
30	Offer alterr	2.943131	3.056869	0.113738
31	Wide range	2.94525	3.05475	0.109501
32				
33	Parameter	Output D41	Output D41	Output Differences:
34	Financial a	2.4807	3.5193	1.0386



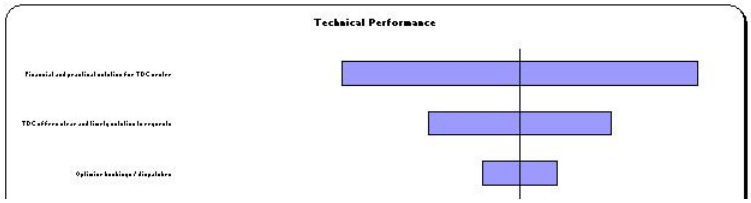


Figure 36: Tornado plots excerpt. The tool also produces a table with the calculated values needed for the chart

6.2 Conclusions on the evaluation tool and further uses

The value of this evaluation tool goes far beyond the computed results. Within a large project, where information is generally scattered among several documents (and people) the tool provides means to aggregate, in a very efficient way, information transversal to all project areas, from technical data to management strategies. The simplicity of these sensitivity analysis procedures allow the user to focus on the most important aspects of a project. In the same line of thought the tool can be easily used to monitor real-time data, providing information for different levels of stakeholders.

The tool was developed using spreadsheet software, as a prototype. For a continuous use on the field, a more error prone solution should be considered.

7 Conclusions and future developments

Main conclusions

The evaluation plan was executed in the five mentioned European locations. The evaluation framework was successfully applied to three pilot projects and two feasibility projects for DRT services. The framework and the evaluation plan were generally well accepted by the partners, although it should be mentioned that the FEUP team made a special effort in explaining the concept to each partner individually, assisting them in the evaluation process.

The results produced by the evaluation tool were validated by the local teams. The use of the tool had a decision support value even in the set-up phases. It helped local teams in tuning both objectives and indicators, resulting in a better correspondence between the actual scenarios and the data representation.

The results also clearly indicated areas of improvement and of concern. This characteristic also helped in the (re)design of feasibility concepts.

Further real use of the tool is needed to actually verify its usefulness as a monitoring tool or as a decision support system, albeit these possible uses seem natural.

The software implementation has some limitations. The tool must be pre-configured by an expert to match the number of objectives of each site. The objective titles can be changed without an expert intervention, but adding or removing objectives must be done with assistance. During the MASCARA project, two sites requested a change in the number of objectives. This does not happen with the number of indicators, although there is a maximum allowed per objective. This maximum can be changed (increased), but not without expert intervention.

Other types of transport services evaluation

The evaluation framework can be easily applied to other types of transport services. In fact, the demand responsive aspect of DRT is reflected only at the indicator level, and even here, just a few set of indicators are solely related to DRT, mainly the ones regarding the Travel Dispatch Centre. Otherwise DRT services share a large common ground with other types of transportation services and most of the work can be reused.

Generic evaluation use

The proposed evaluation framework was designed with the specific objective to evaluate Demand Responsive Transport services. We verified that our evaluation needs fit well a multi-criteria evaluation problem, that can be abstracted from the concrete transport service situation. In fact, the methodology and the tool can be

easily adapted to other cases, requiring a similar structure of the evaluation needs. At this time, the evaluation tool has already been adapted and is being used to evaluate video-games at a software company, also under three main criteria, code (programming), graphics and gameplay which are then sub-divided in several different areas (otherwise objectives) each one ending in a set of metrics (Indicators).

One scenario situations

In most evaluation processes, people have a clear understanding on what to compare the measures taken, generally well established values. The evaluation can then provide a measure against those values, and it somewhat easy to infer an absolute evaluation. In our case we had already a fair number of similar services for which data was also available. Our main problem was that these services were quite different in nature, moreover they are provided in quite different social, geographical, political and cultural environments, which made almost meaningless any comparison.

Relative evaluations are always possible, but, at the very least, they need two scenarios to compare and decide which one is the best, under some criteria. By using the TOPSIS methodology concept of "ideal" and "anti-ideal" values representing an expert opinion, on the range that a particular measure can be expected to vary, we automatically create an ideal and anti-ideal scenario related to the one being evaluated. Having these two virtual scenarios as references, we can easily provide an absolute evaluation. This concept is quite simple, and is probably present in the mind of managers and service designers when setting goals to achieve. Our method has the advantage of explicitly address this kind of intuition, and in doing so, making the evaluation process more clear and understandable.

Simplicity

The proposed methodology, uses rather simple concepts in its base. This is already a recognised advantage of the AHP and TOPSIS methodologies. This aspect becomes more important when the evaluation is public or known to others. The evaluation process must be clear and easily understandable to a wide range of stakeholders, which can be from different cultures and backgrounds.

Monitoring

The evaluation tool was designed to be used once; to evaluate a service concept or an actual running service, but its use can be extended. By repeating the indicators measures in a timely fashion, a service manager can have a "live" representation of the service with several levels of detail. Also, recurring to the sensitivity analysis performed, the managers can decide to focus on a smaller set of indicators, i.e. the most sensitive ones.

Decision Support System

The tool can also be used as a decision support system when used to analyse "what-if" situations. Given a new scenario possibility, the corresponding changes can be entered to the system and new evaluation scores computed. The sensitivity analysis can also provide a good indication on what areas to improve or to focus attention on. A simple example is the evaluation of a new bus purchase. The previewed indicator changes would be entered into the system and the effect measured by the changes in the results.

Future software developments

The implemented tool, using a spreadsheet software, is of a prototype nature. It is relatively low demanding on users, mostly requiring the repeated treatment of information of the same kind (indicators). It is therefore easy to envision more practical ways to gather this information. The tool could be adapted to have a client-server architecture, where the client side could access the server side via a web browser in a multi-user environment. Then it would be possible to give different access to different areas based on the user, e.g some technical users could be responsible for just a set of indicators, and managers could be allowed to enter new objectives. This multi-user environment could be of great advantage especially in a rather distributed business like transportation, which covers several areas, specially in the technical ground.

References

[1] Brake, J., Nelson, J.D. & Wright, S. (2004). Demand responsive transport: towards the emergence of a new market segment. *Journal of Transport Geography*, **12**, pp. 323-337.

[2] http://www.drtbus.co.uk/new%20what%20is%20drt.htm

[3] Brake, J., Mulley, C. & Nelson, J. (2006). *Good practice guide for DRTs using telematics*. Newcastle upon Tyne. Newcastle

[4] Ward, D. (1975). *Theoretical comparison of fixed route bus and flexible route subscription bus service in low density areas*. ed. Transportation Systems Center USDOT.

[5] http://www.scotland.co.uk/Publications/2006/05/22101418/6

[6] Arrillaga, B. & Medville, D. (1974). Demand, supply and cost modelling framework for demand-responsive transportation systems. *Transportation Research Board Special Report*, **147**, pp. 32-47.

[7] Flusberg, M. & Wilson, N. (1976). A descriptive supply model for demand responsive transportation system planning. *Proceedings of the 17th Annual Meeting of the Transportation Research Forum*, , pp. 425-431.

[8] Lerman, S. & Wilson, N. (1974). Analytic model for predicting dial-a-ride system performance. *Transportation Research Board Special Report*, **147**, pp. 48-53.

[9] Daganzo, C. (1978). An approximate analytic model of many-to-many demand responsive transportation systems. *Transportation Research*, **12**, pp. 325-333.

[10] Diana, M., Dessouky, M. & Xia, N. (2006). A model for the fleet sizing of demand responsive transportation services with time windows. *Transportation Research*, **40**, pp. 651-666.

[11] Palmer, K., Dessouky, M. & Abdelmaguid, T. (2004). Impacts of management practices and advanced technologies on demand responsive transit systems. *Transportation Research, Part A: Policy and Practice*, **38**, pp. 495-509.

[12] Quadrifoglio, L., Dessouky & Ordóñez, F. (2005). A Simulation Study of Demand Responsive Transit system Design.. *Submitted for publication*, , pp. .

[13] Savelsbergh, M. & Sol, M. (1995). The general pickup and delivery problem. *Transportation Science*, **29**, pp. 17-29.

[14] Desaulniers, G., Desrosiers, J., Erdmann, A., Solomon, M. & Soumis, F. (2001). *The VRP with pickup and delivery. The Vehicle Routing Problem*. ed. P. Toth and D. Vigo SIAM Monographs on Discrete Mathematics and Applications 9. Philadelphia

[15] Cubillos, C., Guidi-Polanco, F. & Demartini, C. (2004). Multi-Agent Infrastructure for Distributed Planning of Demand-Responsive Passenger Transportation Service. *IEEE International Conference on Systems, Man and Cybernetics*

[16] Hickman, M. & Blume, K. (2000). A method for scheduling integrated transit service. *8th International Conference on Computer-Aided Scheduling of Public Transport (CASPT) Berlin, Germany.*

[17] Horn, M. (2002). Multi-modal and demand-responsive passenger transport systems: a modeling framework with embedded control systems. *Transportation Research*, **36A** , pp. 167-188.

[18] Horn, M. (2002). Fleet scheduling and dispatching for demand-responsive passenger services. *Transportation Research*, **10C**, pp. 35-63.

[19] Horn, M. (2003). An extended model and procedural framework for planning multi-modal passenger journeys. *Transportation Research*, **37 Part B** , pp. 641–660.

[20] Mageean, J. (2003). Mageean, J, Nelson, J.D. - The evaluation of demand responsive transport services in Europe, Journal of Transport Geography 11 (2003), pp. 255-270. , , pp. .

[21] Stone, J., Nalevanko, A. & Gilbert, G. (1994). Computer dispatch and scheduling for paratransit: an application of advanced public transportation systems. *Transportation Quarterly*, **48**, pp. 173-184.

[22] Goeddel, D. (1996). *Benefits assessment of advanced public transportation systems*. ed. U. S. Department of Transportation . Washington, D.C.

[23] Ben-Akiva, M., Bejamnin, J., Lauprete, J. & Polydoropoulou, A. (1996). Impact of advanced public transportation systems on travel by dial-a-ride. *Transportation Research Record*, **1557**, pp. 72-79.

[24] Chira-Chavala, R. & Venter, C. (1997). Cost and productivity impacts of a smart paratransit system. *Transportation Research Record*, **1571**, pp. 81-87.

[25] Wallace, R. (1997). Paratransit customer: modeling elements of satisfaction with service. *Transportation Research Record*, **1571**, pp. 59-66.

[26] Higgins, L., Laughlin, J. & Turnbull, K. (2000). Automatic vehicle location and advanced paratransit scheduling at Houston METROLift. *Proceedings of the 2000 Transportation Research Board Conference, Washington, D.C.*

[27] Stone, J., Ahmed, T. & Nalevanko, A. (2000). Internet-based decision support for advanced public transportation systems technology. *Proceedings of the 2000 Transportation Research Board Conference,Washington, D.C.*

[28] Valenti, G., Perugia, A., Binazzi, C., Di Volo, N. & Ambrosino, G. (2001). *The role of the Demand Responsive Transport Services in sustainable mobility: the experience in Florence*. ITS. Prague

[29] Moriyama, M., Fujiwara, A. & Zhang, J. (2005). Development of non-monetary indicators to evaluate public transport services in rural areas. *Journal of the Eastern Asia Society for Transportation Studies*, **6**, pp. 423 - 434.

[30] Halden, D., TAS Partnership & University of Aberdeen (2006). *How to plan and run Flexible and Demand Responsive Transport*. Transport Research Planning Group, Social Research,Scottish Executive. Edinburgh

[31] European Conference of Ministers of Transport (May 2006). *Improving Transport Accessibility for All: Guide to Good Practice Complete Edition*. . ISBN 9282101398

[32] Halden, D., the TAS Partnership & University of Aberdeen (2006). *Review of Demand Responsive Transport in Scotland* . .

[33] Valiris, G., Chytas, P. & Glykas, M. (2005). Making decisions using the balanced

scorecard and the simple multi-attribute rating technique. *Performance Measurement and Metrics*, **6**, pp. 159-171.

[34] Driva, H., Pawar, K. & Menon, U. (2001). Performance evaluation of new product development from a company perspective. *Integrated Manufacturing Systems*, **12, No. 5**, pp. 368-378.

[35] Saad, G. (2001). Strategic performance evaluation: descriptive and prescriptive analysis. *INDUSTRIAL MANAGEMENT & DATA SYSTEMS*, **101**, pp. 390-399.

[36] Saaty, T. (1980). *The Analytic Hierarchy Process*. McGraw-Hill.

[37] Brans, J. & Vincke, P. (1985). A preference ranking organization method. *Management Science*, **31 No. 4**, pp. 647-656.

[38] Brans, J., Vincke, P. & Marechal, B. (1986). How to select and how to rank projects - The PROMETHEE method. *European journal of operational research*, **24**, pp. 228-238.

[39] Figueira, J., Greco, S. & Ehrgott, M. (2005). *Multiple criteria decision analisys: state of the art surveys*. Kluwer Academic Publishers.

[40] Rao, R. & Davim, J. (2008). A decision-making framework model for material selection using a combined multiple attribute decision-making method. *International journal of advanced manufacturing technology*, **35**, pp. 751-760.

[41] Water, H. & Vries, J. (2006). Choosing a quality improvement project using the analytic hierarchy process. *International Journal of Quality & Reliability Management*, **23, Num. 4**, pp. 409-425.

[42] Yurdakul, M. & Ic, Y. (2005). Development of a performance measurement model for manufacturing companies using the AHP and TOPSIS approaches. *International journal of production research*, **43**, pp. 4609-4641.

[43] Yurdakul, M. (2002). Measuring a manufacturing system's performance using Saaty's system with feedback approach. *Integrated Manufacturing Systems*, **13, No. 1**, pp. 25-34.

[44] Gass, S. & Rapcsak, T. (2004). Singular value decomposition in AHP. *European journal of operational research*, **154**, pp. 573-584.

[45] Saaty, T. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary. *European journal of operational research*, **145**, pp. 85-91.

[46] Kwiesielewicz, M. & van Uden, E. (2004). Inconsistent and contradictory judgements in pairwise comparison method in the AHP. *Computers & operational research*, **31**, pp. 713-719.

[47] Buede, D. & Maxwell, D. (1995). Rank disagreement: A comparison of multi-criteria methodologies. *Journal of Multi-Criteria Decision Analysis*, **4**, pp. 1-21.

[48] Pöyhonen, M., Hämäläinen, R. & Salo, A. (1997). An experiment on the Numerical Modeling of Verbal Ratio Statments. *Journal of Multi-Criteria Decision Analysis*, **6, issue 1**, pp. 1-10.

[49] Pérez, J., Jimeno, J. & Mokotoff, E. (2006). Another Potencial shortcoming of AHP. *TOP*, **11, N.1**, pp. 99-111.

[50] Benayoun, R., Roy, B. & Sussman, N. (1966). Manual de Reference du Programme

Electre. *Note de Synthese et Formation,25*, , pp. .

[51] Wang, X. & Triantaphyllou, E. (2004). Some ranking irregularities when the ELECTRE method is used for decision-making. *presented at the Ind. Eng. Res. Conf. (IERC) Houston, TX*

[52] Janko, W. & Bernroider, E. (2005). Multi-Criteria Decision Making: an application study of Electre & Topsis. *gennaio*, , pp. 39,40.

[53] Matos, M. (2004). Multicriteria techniques. *Lecture Notes* - Escola de Gestão do Porto

[54] Munda, J. (2006). A NAIADE based approach for sustainability benchmarking . *International Journal of Environmental Technology and Management*, **6, Num.1-2**, pp. 65-78.

[55] Hwang, J. & Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. ed. Springer Verlag . New York

[56] Yurdakul, M. & Cogun, C. (2003). Development of a multi-attribute selection procedure for non-traditional machining processes. *Proceedings of the institution of mechanical engineers part b-journal of engineering manufacture*, **217**, pp. 993-1009.

[57] Zanakis, S., Solomon, A., Wishart, N. & Dublish, S. (1998). Multi-attribute decision making: A simulation comparison of select methods. *European journal of operational research*, **107**, pp. 507-529.

[58] Isiklar, G. & Buyukozkan, G. (2007). Using a multi-criteria decision making approach to evaluate mobile phone alternatives. *Computer standarts & interfaces*, **29**, pp. 265-274.

[59] Majumdar, A., Sarkar, B. & Majumdar, P. (2005). Determination of quality value of cotton fibre using hybrid AHP-TOPSIS method of multi-criteria decision-making. *Textile institute*, **96, Nº5**, pp. 303-310.

[60] Lin, M., Wang, C., Chen, M. & Chang, C. (2008). Using AHP and TOPSIS approaches in customer-driven product design process. *Computers in industry*, **59**, pp. 17-31.

[61] Yang, T., Chen, M. & Hung, C. (2007). Multiple attribute decision-making methods for the dynamic operator allocation problem. *Mathematics and computers in simulation*, **73**, pp. 285-299.

[62] Tzeng, G., Lin, C. & Opricovic, S. (2005). Multi-criteria analysis of alternative-fuel buses for public transportation. *Energy policy*, **33**, pp. 1373-1383.

[63] Miller, G. (1994). The magical number 7, plus or minus 2 - some limits on our capacity for processing information (reprinted from psychological review, vol 63, pg 81, 1956). *Psychological review*, **101**, pp. 343-352.

[64] Belton, V. & Stewart, T. (2002). *Multi Criteria Decision Analysis – An Integrated Approach*. ed. Kluwer .

[65] Eschenbach, T. (2006). Technical note: constructing tornado diagrams with spreadsheets. *The Engineering Economist*, **51**, pp. 195-204.