DXX-GTF Recommendations & Dissemination activities

Project: SPOTLIGHTS-TN (Thematic Network)

Contract No 1999-TN.10941

Project Co–ordinator

MCRIT SL (Barcelona)

WP Leader

MKmetric Gesellschaft für Systemplanung mbH (Karlsruhe, Germany)

WP Contributor

-

Other Partners

DTU, Technical University of Denmark (Copenhagen, Denmark)

Date: 5th February 2002

PROJECT FUNDED BY THE EUROPEAN
COMMISSION UNDER THE TRANSPORT
RTD PROGRAMME OF THE
5th FRAMEWORK PROGRAMME

Responsibilities & Contacts:

European Commission Responsible

Commission of the European Communities
Directorate–Transport and Energy (TREN)
Anna Panagopoulou
Rue de la Loi 200
B –1049 Brussels
Belgium

Tel: +32 (2) 296 78 94
Fax: +32 (2) 295 43 49
email: anna.panagopoulou@cec.eu.int

WP Leader

MKmetric Gesellschaft für Systemplanung mbH
Durlacher Allee 49
D – 76131 Karlsruhe
Germany

Tel.: +49 721 961 60 0
Fax.: +49 721 961 60 99

Theory:
Dr. Benedikt Mandel
email: mandel@mkm.de

Project & Development:
Eduard Ruffert
email: ruffert@mkm.de

Contents

1 RECOMMENDATIONS

1.1 REQUEST FOR A COMMON UNDERSTANDING TOWARDS GTF

1.1.1 WHAT IS THE "GENERALISED TRANSPORTATION-DATA FORMAT" (GTF)?

GTF, currently, is a specification of a conceptual model (GTF-CM) structuring the terms, language and concepts found in the problem domain of transportation modelling, esp. strategic, by formal definition. As a result these terms, language and concepts can be understood and processed automatically by computers, see the following figure.

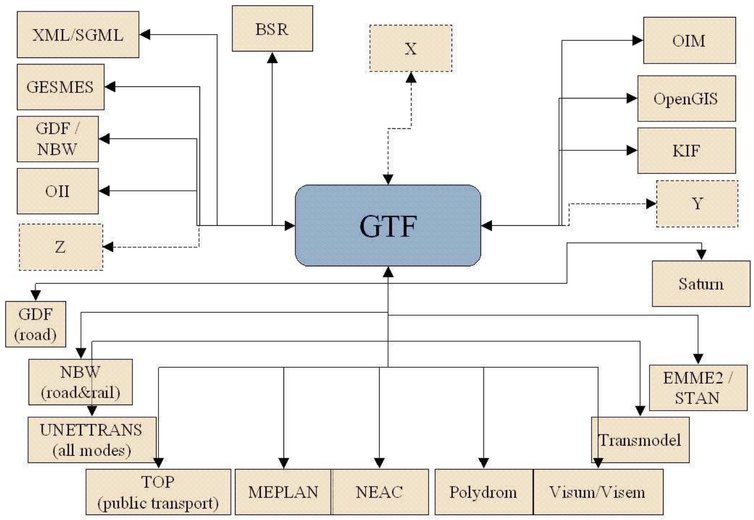
GTF is the first effort to solve the problem of "communication" between different transportation models and between transportation models and other software or computer systems. Currently, "Communication" takes place only on the basis of exchanging "data". But this is not enough as the projects from the 4th and 5th Framework Programmes have analysed. What is needed, is an exchange of "information", which is defined as "data" plus "meaning". Data structured according to the GTF-CM specification adds the necessary "meaning" to the raw "data". (An effort in this direction is NESSTAR which is funded by DGXIII of the European Commission under the 4th Framework Telematics Applications Programme. http://www.nesstar.org. It is based on the DDI DTD, a standard proposed by an international committee of data librarians and statisticians. The consortium has already, and is continuing to, put together a range of data and metadata loaders for this standard.)

Only data with meaning, i.e. information, has the necessary semantics required for meaningful further usage of the data. For example, data that is the output from a transportation model showing the impacts of a policy measure on the flows of goods and passengers between some regions can only be used by a subsequent other transportation model, e.g. computing the ecological effects of the changes in passenger and goods flows, if not only the raw data is transferred to the other model (- this is easy to achieve) but also the meaning and exact definitions of each output column of data from the first model matches exactly the meanings and definitions from the second model.

This interoperability which gives a value added to the outputs from the transportation models is not possible without the concordance, otherwise problems arise like inconsistency of the results or the results cannot reasonably be explained or are misunderstood completely. The outcome from this is that model results cannot be reused for additional purposes or problem analysis.

The “GTF vision” is to reach a level of interoperability of transportation models through harmonisation of their input/output structures and protocols, i.e. by providing input/output software “translators” that convert from proprietary data structures to and from GTF (mediator) structures.

This would dramatically reduce the number of translators (to 2*n) that would be needed in contrast to bilateral (point-to-point) translation (from n^2), which is the current “state-of-the-art”, see the following figure. This is because the multiplicative effect of adding another transportation model format is reduced to a linear effect by the central mediator function of the GTF format.



This concept has the advantage that databases and transportation models’ software will not need to be completely restructured to comply with GTF. The only requirement will be that the computer system hosting the database and transportation models must provide GTF translators.

1.1.2 GOALS AND BENEFITS OF GTF

1) Interoperability of transportation models, i.e. results from one model could be used more easily as input for another model in an automatic fashion

2) Competition between transportation models increasing the quality of results and the transparency of procedures and methods of transportation models, i.e. they will not be "black-boxes"

3) Plausibility and understanding of model results due to possibility of comparison of different models' results

4) Re-use of data, therefore more efficient use of resources and committed public funds

5) GTF is an additional opportunity and incentive for harmonisation of transportation models input/output and data on the administrative / managerial level (not just the technical level)

6) "Club idea" of GTF users as concrete step to solve licensing and legal issues in a structured manner (and therefore cost and time reduction to achieve the same)

7) Development of a "GTF Translator Toolkit" (a library of functions) to help, drive, concentrate and guide the efforts of GTF Translator development

8) GTF as "glue" for the concepts proposed for ETIS

9) Consistency and transparency of decision-making processes that involve the use of transportation models

10) Use of a sample or collaboration the "best" models for problem solving and decision making

11) Reduction of "risk" of non-optimal decisions due to comparison of multiple results from different models. Cost reduction by avoiding non-optimal decisions due to comparison of multiple results from different models. Quick and efficient solving of complex queries (for decision making) because of availability of several transportation models with easy access

12) Simplification of the use of transportation models and data for problem solving. GTF is an elementary and necessary step to support and enhance the work at the decision-making level

13) Clear benefits are to be gained for the public sector (short- and mid-term) as well as for the research field and the private sector (mid- and long-term) by supporting complex question problem solving which nowadays cannot be solved efficiently due to missing information (data) and missing tools (software and system) and non-interoperability. GTF cannot solve the problem of really missing data, it can help in making the data used in different models more easily available and understandable.

14) GTF is the consequent next step to uptake the results in the field of transportation modelling from the 4th and 5th Framework

15) Better transparency, acceptance and understanding of political decisions and the decision-process by the citizens

16) The research field in transportation models will be boosted, encouraged and improved

17) Integration and concentration of plurality of opinions through the use of a standard discussion basis, e.g. the GTF Forum or the Mailinglist

18) GTF as connection / key to international exchange of experiences and knowledge in the transportation field on the level of the decision-making tools (transportation models and data)

19) Improvement of the link between databases and transport models by an effort to make transport databases more compatible in their content and structure with the requirements of modelling (which could lead to a standardisation and a better interpretation of transport data produced by institutions and commercial data providers). The adoption of the GTF conceptual model for transport databases would make it much easier to analyse and understand their content and to compare data among different sources.

1.1.3 Recommendations based on the spotlightsTN GTF workshop

According to the experts of the GTF workshop the benefits and gains that could be reached (as mentioned in the previous section) can be attained by endorsing and seriously supporting the following recommendations. The experts of the spotlightsTN/GTF Workshop held on 11th / 12th October 2001 in Barcelona hereby recommend the following.

1) The European Commission should endorse the finalisation of the “Generalised Transportation-data Format” Specification (GTF v1.0) while also moving towards the implementation level (see 7-10) in the context of the European Transport policy Information System (ETIS) development projects.
2) The European Commission should endorse to make available its own data and data from any projects and studies undertaken (using the developed GTF Translators) in the GTF format thus reducing costs for future projects by re-use of data and information from previous projects.
3) The European Commission should recommend to its institutions and the member states to support the finalisation of the GTF Specification and to take steps to provide data in the GTF format (using the developed GTF Translators). Please refer to the results of the ATOM project.
4) To streamline, organise and optimise the introduction of the “Generalised Transportation-data Format” the European Commission should think about establishing an independent advisory board consisting of scientific, institutional and industrial members. Subsequently this board would work out guidelines for the procedure to establish GTF and would deal with and resolve all administrative and legal matters as well as all questions concerning data provision and transmission rules, issues on the continuing maintenance and updating of GTF etc.
5) The European Commission should consider fostering the world wide dissemination of knowledge on GTF at transport and software development conferences as well as publishing papers on GTF -in detail and generally about the GTF vision- in appropriate and / or relevant scientific and software development journals.
6) The GTF Forum (http://gtf.mkm.de) or Mailinglist (gtf@lists.mkm.de) should be kept open as the permanent and standard means of knowledge exchange on all issues of the GTF task, so that the evolution of GTF can take place under consideration of societal and technical development and the open participation of all interested parties.
7) From “conceptual model” to ”data model”, i.e. the GTF-CM has to reach the next phase which would be the design and implementation of a specific data model (which can be used in actual databases) thus making the complete GTF-CM operational, not only parts as implemented in current systems, e.g. ATMax or TOP.
8) From “common agreement” to ”ontology”, i.e. the ”common understanding” could effectively be used as a basis to develop a glossary (mid term) and then an ontology (long term) of knowledge in the problem domain of transportation modelling thus reducing costs and loss of time due to inconsistencies in definitions of data and transportation model outputs.

9) "Library of functions", i.e. a necessary consequence of up-taking the results from the 4th and 5th Framework Program projects would be the development of a freely available library of functions e.g. reading/writing functions for GTF files, which can be used by potential software developers of GTF Translators, thus reducing their costs and reducing the "time-to-market" of GTF Translators, making the functionality of "interoperability of models" available to decision makers faster. The library should be accessible to any person, organisation or company to build a complete GTF Translator to the own proprietary software data structures. Therefore the European Commission should instigate a pilot project to prove further the GTF concept in practice. One of the outputs of this would be the foundations of the "library of functions".

10) "Standardisation", i.e. after further work has been done on GTF it should be evaluated by the European Commission whether to submit the GTF Specification to a standardisation board to make it a standard (at least European wide), e.g. the European Committee for Standardisation CEN (http://www.cenorm.be/), the International Organisation for Standardisation ISO (http://www.iso.ch) or any other board deemed appropriate.

1.1.4 EXAMPLE DISCUSSION ON RECOMMENDATIONS DOCUMENT

Comments from Michel Houée:

To start with, I would like to state that I clearly share with the authors of the memorandum of understanding towards a “generalised Transportation-data Format” (GTF) the idea that any effort in the direction of making the transportation models more interoperable and their results more comparable is certainly worth spending, considering the difficulty experienced in many circumstances within the European research projects as well as within the framework of bilateral or multilateral cooperation to explain the differences in the outcomes of models. The last Friday workshop of the Think-up project is a good illustration of this difficulty, nobody in the audience, nor the national delegates neither the coordinators of the Expedite project, being in a position to explain why the reprocessing by Expedite of the Scenes model led to such a larger elasticity of rail freight traffic to fare and time as compared with the rather converging features of the national models to that respect.

The question that is raised is therefore : what could be the contribution of a GTF to such a perspective.

The feeling I get from the reading of the last version of this memorandum (now called request) is that it would help in order to identify this contribution to split between the specific core of GTF and the complementary initiatives that are requested to make it efficient.

My own understanding is that this specific core is the set of input/output “translators” that would “convert from proprietary data structures to and from GTF (mediator) structure”.

The peripheral complementary initiatives requested are mainly in my mind :

- the full documentation of the diverse transport data sources that are likely to be used in the various types of models applied in the transportation sector, resulting optimally in a library of metadata.
- the full documentation of the properties of each model considered, which covers a wide range of areas, from the mathematical structure of the model up to the type of geographical zoning it admits for instance.

I do not personally believe that the two last items can be achieved within a single project, whatever it is called GTF or anything else, even if the need of this achievement for implementing the translators might be an incentive among others in that direction.

Documentation of data is certainly a process that has already been at work for long within official institutions such as Eurostat as well as at the spontaneous initiative of some of the data providers, although in my opinion a great deal remains to be done.

Documentation of models is probably even more embryonic, due to the greater complexity of the description as well as to the reluctancy of model owners to display their “building secrets” to possible competitors or reveal the possible weaknesses of their tool.

That means in my mind that progress can only be achieved if :

- beyond official statistics the field of which will necessarily remain limited (and the extension of which may take years of negotiation with Member States), a significant change of behaviour is produced among professionnal data owners concerning the benefit they may expect from a mutual exchange of data made available to third parties;
- a strong incentive is given by European institutions to more transparency in model characteristics, for instance as a pre-request for labelling these models as of European interest.

In addition, considering the data side of the problem, it is not enough to provide full documentation of data sets to make sure they can be used in a comparable way. Quite often, referring to a common field in two different surveys (same definition of the population in terms of age threshold, or of trips considered in terms of distance threshold for instance) is only possible at the condition of a reprocessing of one of the two surveys or by the adoption of possibly very hypothetic assumptions. And it is a very long and uncertain process indeed that may lead, through statistical directives and regulations as well as through specific projects such as Dateline, to a real homogeneity of data at European level.

This do not mean that the translators cannot be an important element of the whole, which as such deserves some support. But this support would certainly be more easily gained if some demonstration of their capability was evidenced from previous experiences. From that point of view, it could be appropriate to elicit in the memorandum examples of the kind mentioned by André Ulied during last Spotlights meeting in Barcelona concerning projects done on behalf of Barcelona authorities; and if similar examples could be derived from Atom, it would even reinforce the coherence of the whole.

As a consequence of these considerations, I would make the following suggestions/comments concerning the “memorandum” :

2. May be the chapter 2 would be clarified by splitting between goals and benefits.

2.1 I do not trust that interoperability of models can be produced “in an automatic fashion” I would rather state “results from one model could be used more easily as input for another model”.

2.2 I see more transparency in models as a pre-requisite to fully benefit from GTF than as a consequence of GTF (see suggestion about 2.).

2.5 Presenting GTF as a “first step” for harmonisation of data, even restricting to the administrative/managerial level, is not correct, it would be better stated as “an additional opportunity and incentive to make progress” in that direction.

2.13 GTF as such cannot solve the problems due to really missing information (data), it can only (which is a lot indeed) help in making data used in different models more easily available and understandable.

2.14 Probably less ambiguous if it is specified “results in the field of transportation modelling of the 4th and 5th FP”.

2.19 Shouldn’t it be recognised that models are not the only possible way of using data by speaking, instead of “standardisation”, of effort to make transport databases more compatible in their content and structure with the requirements of modelling.

3.3 It is probably necessary in order to get the support of Member States in the finalisation the GTF Specification to organise, beyond the reference to some demonstration cases of GTF application in the memorandum of understanding, a debate about the usefulness and, most important, feasibility of the GTF, why not through a final workshop of dissemination of the Spotlights results, may be together with the dissemination of the ATOM results and possibly of other related projects.
The question of providing (national) data in the GTF format using the GTF translators would obviously require in a second step a more specific investigation of the compatibility problems that may occur for certain types of data.

3.7 and 3.8 It is not totally clear to me what is meant exactly by “data model” (just the idea of making GTF effective, or something more precise), and by ‘ontology” in concrete terms.

3.9 I totally support the idea that “the European Commission should instigate a pilot project to prove further the GTF concept in practice”

Reply by MKmetric

Dear Michel.

Thank you for your comments and your opinions.
We will incorporate many of the formulations you suggest below.

But we have to stress one or two points for clarification:

’The question that is raised is therefore : what could be the contribution of a GTF to such a perspective.

My own understanding is that this specific core is the set of input/output “translators” that would “convert from proprietary data structures to and from GTF (mediator) structure”.'

This is correct but the following conclusions we cannot share

'The peripheral complementary initiatives requested are mainly in my mind :
- the full documentation of the diverse transport data sources that are likely to be used in the various types of models applied in the transportation sector, resulting optimally in a library of metadata.
- the full documentation of the properties of each model considered, which covers a wide range of areas, from the mathematical structure of the model up to the type of geographical zoning it admits for instance. '

because these tasks were explicitly not part of the GTF. E.g. The "full documentation of data types" was a task of spotlights/SPQR (and originally was the task of BRIDGES/DDG Digital Data sources Guide). The "full documentation of properties of models" was the work in spotlights/Mdir + spotlights/SPQR.

Do not misunderstand us, we share your view that these things have to be addressed and done, but it was not part of what we (MKmetric) were supposed to do in GTF.

Therefore this too we have to point out

'I do not personally believe that the two last items can be achieved within a single project, whatever it is called GTF or anything else, even if the need of this achievement for implementing the translators might be an incentive among others in that direction.

Documentation of data is certainly a process that has already been at work for long within official institutions such as Eurostat as well as at the spontaneous initiative of some of the data providers, although in my opinion a great deal remains to be done.

Documentation of models is probably even more embryonic, due to the greater complexity of the description as well as to the reluctancy of model owners to display their “building secrets” to possible competitors or reveal the possible weaknesses of their tool. '

is was not part of the GTF works. (It was explicitly stated in the contract that model documentation etc. are tasks of SPQR and Mdir.)

We agree with the rest of the points you raise and will incorporate them in the final version of the "Recommendations" document.

One last thing concerning your comment on the feasibility or usefulness of GTF. Here, ATOM in its D3 (yet to be approved) show that GTF is in comparison a more useful concept than e.g. UNETRANS or GDF. And they made an application of translating some SCENES data into GTF-XML (GTF-XML is the concrete data format of the GTF-CM).

They came to the conclusion that it is feasible and easier than to translate into other formats e.g. UNETRANS, GDF ...

Thank you for your detailed and insightful comments and opinions.

1.1.5 Statement of the undersigning institute(s)

The signatories endorse and support GTF, the “GTF vision”, deem it very necessary, acknowledge the benefits and fully agree to the recommendations.

Previous version of the recommendations signed by:

Institution/Company	Name
ETH Zurich	Prof Nagel
ETH Zurich	Prof Axhausen
Polydrom	Mr de Rham
INRO	Prof Florian
Nestear	Mr Reynaud
Steer Davies Group	Dr Willumsen

Letter of Intent signed by

Mr Reynaud (Nestear)

Nestear
Nouveaux ESpaces de Transport en Europe
Applications de Recherche

Adresse des bureaux (Office address)
89-93, Av. Paul Vaillant Couturier
94250 GENTILLY

Tél :33 1 41 98 38 10
Fax : 33 1 45 46 55 12

Gentilly, 2 Janvier 2002

Letter of intent concerning GTF

We hereby endorse and support GTF and the "GTF vision", deem it very necessary, acknowledge the benefits and fully agree to the recommendations as described in the "Recommendations to the European Commission" document.

The Director of NESTEAR

Christian REYNAUD

Siège social : 99bis, Avenue du Général Leclerc
75014 PARIS
R.C.S. PARIS B 437 573 207
http://www.neste.net

1.2 LETTER OF INTENT

Letter-of-Intent concerning GTF

We hereby endorse and support GTF and the “GTF vision”, deem it very necessary, acknowledge the benefits and fully agree to the recommendations as described in the “Recommendations to the European Commission” document.

Signatories,

2 DISSEMINATION ACTIVITIES

2.1 PAPER SUBMITTED TO PTRC: FOR THE "EUROPEAN TRANSPORT CONFERENCE" ETC, 10-12TH SEPTEMBER 2001

GENERALISED TRANSPORTATION-DATA FORMAT (GTF)

- DATA, MODEL AND MACHINE INTERACTION

Otto Anker Nielsen, Professor, Ph.D.
Centre for Traffic and Transport, Technical University of Denmark
(oan@ctt.dtu.dk)

Dr. Benedikt Mandel and Eduard Ruffert
MKmetric Gesellschaft für Systemplanung, Karlsruhe, Germany
(mandel@mkm.de, ruffert@mkm.de)

2.1.1 INTRODUCTION

Exchanging data and information on the data (meta-data) between transport models, as well as between transport models and other software, e.g. GIS, is always a very tedious, if even possible, task. There is often the problem of loss of information because the exchanged data only seemingly contains the information required. And there is also often the problem of inhomogeneous and proprietary data formats forcing the users of the data to re-format and re-combine the data from scratch every time.

This is both due to 'low-level' differences in data formats, and due to more fundamental 'high-level' differences in the conceptual models, e.g. for network topologies. Examples of the latter are the differences in describing a terminal by transfer tables versus by a sub-network, or a public transport network by timetables referring to the same line, versus by parallel arcs for each departure.

The solution to these problems is that not only data needs to be transferred, but also the precise meaning of the data (meta-data), including the underlying conceptual model. The 'Generalised Transportation-data Format' GTF, based on the original work in Mandel & Ruffert E. (1999 & 2000) was developed to meet these demands (Note that the name GTF, especially the 'Format' part, stems from its origin trying to find a common format. This subsequently evolved to a specification of a conceptual model, yet the name GTF was retained).

GTF is a proposed conceptual model (covering the most widely used objects in transport modelling), an exchange format (GTF-XML) based on standard XML, and an interchange language to run transport models and retrieve results. This allows software applications, 'GTF Translators', to exchange information and data between transport models and other software.

The work started in the EU-research project BRIDGES where a survey of different conceptual models and formats was carried out (Nielsen *et al*, 1998). This lead to the first version of GTF (Mandel & Ruffert, 1999). The work is continued and refined in the thematic network: SPOTLIGHTS under EU's 5th framework programme, where further surveys, reviews and user input are carried out. This includes co-ordination with the Transport Object Platform, TOP, (Nielsen *et al* 2001a - c), and experiences from the GIS-world, including the US-funded UNE-TRANS-consortia.

As SPOTLIGHTS is funded by the EU, it is the ambition that GTF eventually will become a EU-standard for the exchange of transport modelling data. This will provide a strong platform for utilising earlier work and transport models when building new transport models, as well as a tool for comparing transport models that cover the same geographic area. Both aims will be very useful for research as well as practice in the field of transport modelling.

After an introduction in section 2 to the current situation and problems, the paper suggests an information structure (section 3), entities (section 4) and an exchange format (section 5). The present paper describes GTF in general, while a description of the exchange format and TIP (a protocol to run transport models remotely i.e. through the Internet and retrieve results) would be too voluminous.

To set the work into perspective, comparison with the GIS-based Transport Object Platform (TOP) for public transport is carried out. Finally the conceptual model is summarised and the perspective – and organisational hurdles - for the future use of GTF are outlined in section 7. The paper includes a list of projects and acronyms following the references.

2.1.2 CURRENT SITUATION AND PROBLEMS

The usual use of strategic transport models is to define changes in the input data for each scenario to be analysed. The Input defines ‘Policy Scenarios’, like economic, demographic and spatial developments as well as network changes and changes in prices and fares for the use of transport supply (Eurostat, 1996).

2.1.2.1 Software and transport model issues

Currently, the numerous software applications and databases used in practice are often inhomogeneous and largely incompatible with each other. This leads frequently to problems when comparing results from scenarios based on different software applications and databases.

Transportation modelling at the European level usually requests data from many sources and models – often at least from each country. But often also from different sectors, e.g. road administrations, rail authorities, bus operators, ferry companies, airline systems, etc. This also applies to national models. In some circumstances even further data from non-national sources are needed, e.g. from counties or even municipalities concerning road network data.

As an example, the national road administrations may only maintain databases of the national roads. Since the motorways and highways often end outside harbour cities, the omission of municipal roads can result in large detours in a model. Even some motorways may be owned and maintained by counties, municipalities or private companies.

As such, there are many benefits in integrating data from different sources and at different quality levels. However, figure 1 illustrates the possible problems doing this. This includes:

1. Models will often be inhomogeneous in their conceptual structures, which makes coordination difficult. Furthermore the data models are in some cases not transparent; e.g. software packages are for competitive reasons not fully documented, or they have not been documented properly due to time- and budget constraints.
2. Software packages have inhomogeneous formats (even if they build on similar conceptual models).
3. Some metadata are implicitly given by the software package, and some by the data model; e.g. that an organisation always uses the same unit definition and data collection method. As such the unit definition, quality, year etc. are not stated explicitly in the data itself.
4. Translators are not always sufficient; data may have been aggregated during export, some topological relationships may have been lost during translations, metadata is not exchanged, etc.
5. When data from different sources are combined into one model, there are a number of consistency problems as well as problems stemming from different levels of aggregation.

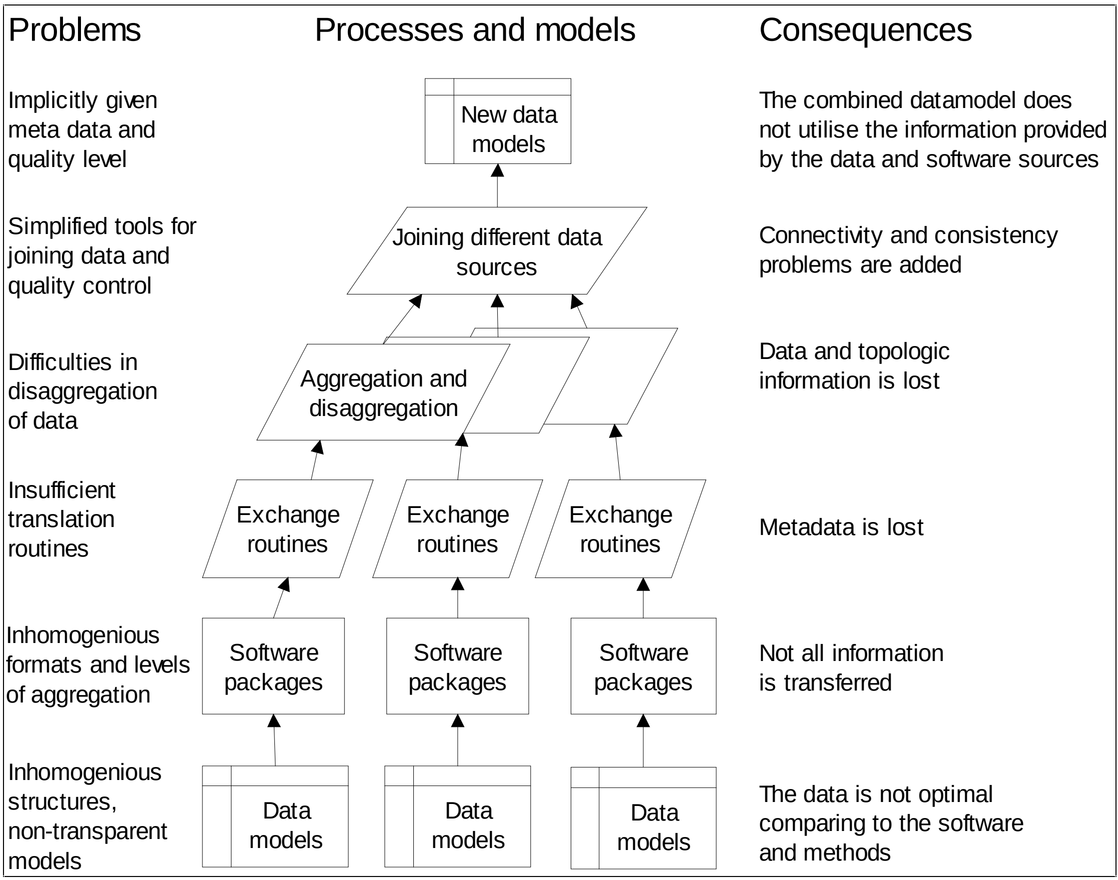


***Figure 1**. Possible problems integrating different models and data sources.*

The problems also apply to the databases of the results from transport models (not only on the input databases), and hereby to the comparison of results from different transport models for different projects, or for the sake of quality control.

2.1.2.2 Problems due to data of transportation models

Even if the above technical problems are solved, problems may still prevail. Transportation models are in general very demanding concerning the amount and quality of input and calibration data. The main problems with current data and databases are:

6. Data required by the model, e.g. for estimation, is not available. For example, a pan-European passenger transport model requires homogeneous input data from all countries.
7. The composition of the available data required by the model does not match and re-composition is not possible. For example, the data acquired for a model has different levels of aggregation or use different segmentation, that cannot be matched to the one needed by the model.
8. The data itself does not match, e.g. that units have been defined differently without an easy way of reformatting this. An example is traffic counts as weekday traffic defined as September to June average, versus traffic counts as Annual Daily Traffic (ADT).

2.1.3 VISIONS BEHIND THE GENERALISED TRANSPORTATION-DATA FORMAT (GTF)

Because of the problems mentioned in section 2, the value of transport models' databases can be significantly increased by homogenising them and by defining an openly available specification of the homogenised conceptual model. The first (and main) advantage would be to have databases, which can be exchanged, enriched, corrected and used in a transparent manner since all would be based on the same conceptual model. Secondly, it can be ensured that the required information is actually contained in the data and that the information can be exchanged. The structure of a 'Generalised Transportation-data Format' accordingly accomplishes the following:

9. Instead of having disparate and manifold software applications and databases, GTF contains all necessary elements and provides one single and homogenous data specification and format.
10. Instead of having incompatible proprietary formats and informational contents, GTF should be used throughout any computer system, by providing translators to / from the proprietary formats to GTF.

To achieve this, GTF consists of:

11. A conceptual model (GTF-CM, called GTF-Conceptual Model). This defines the framework for a given model, while it does not contain the data within the model and the implementation of the model (i.e. it does not constrain any implementation for example as relational tables or as software in anyway).
12. A standard exchange format (GTF-XML), including meta-data as well as the data itself (i.e. 'tags' encapsulating raw data giving it meaning).
13. Generic commands to run models and retrieve results (GTF-TIP, 'Transportation-data Interchange Protocol').

2.1.3.1 Basic concepts

Basically, GTF is a framework, which can be used to define the information that is contained in data. It wraps data into information entities describing the basic data and the necessary supplementary information (meta–data) to give a meaning to the basic data.

A potential problem is, that most models, standard software, and exchange formats define data with implicit information, where only the developer or in the best cases the practitioner with good knowledge of a well-written documentation know the exact definition of a data element, e.g. speed. This needs further definitions, *metadata*, to be defined precisely, e.g.:

14. What *type* of speed; free flow, at congestion, in average, measured, modelled, signed?
15. At what *level of aggregation*; for all lanes, for passenger cars, rush hour, weekday average, all week average, yearly average?
16. *Quality*; measured at each link, judged from road category, guessed on intuition, and method of establishing the data?
17. *Origin*; what is the year of data measurement and updates?
18. *Organisation*; who established the data?

2.1.3.2 XML

In GTF, XML (se e.g. Marchal, 1998 or Booch *et al*, 1999) is used as a framework to ease the definition and exchange of data. The ideas behind XML are a bit similar to those of object-oriented programming (see Brown, 1997, Budd, 1997 or Rumbaugh *et al* for an introduction to object oriented concepts, or to the brief introduction in Nielsen & Frederiksen 2001b). Accordingly, the GTF-XML file includes entity instances and definitions on the relationships between them. The main advantage of this is the minimal amount of different abstract concepts used to cover a wide range of concrete things. The GTF specification defines its exchange format as an application of XML.

2.1.3.3 Main entities in GTF

Very generally speaking, most transportation models use the following information items for their computations (although some models have more advanced input requirements):

19. <u>Zonal data:</u> any kind of zonal description, e.g. socio–economic data, ecological data, zonal boundaries, transport data, indicators, transport matrices etc.
20. <u>Network data:</u> data describing the relations between the elements, e.g. link characteristics, a link has a starting node and an ending node (i.e. topological characteristics), link/network clusters etc.
21. <u>Transport data:</u> data describing services in the public transport, pre-defined routes, etc.
22. <u>GIS data:</u> the necessary information for visualisation purposes, e.g. the underlying projection of the node and its co–ordinates.

The *basic entities* that were used to create the conceptual model are a total of only 10, namely *Node, Link, Mode, Vessel, Chain, DynamicSegmentation, Alternative, Unit, Group* and *Meta*, which are called 'topmost classes' or 'top–levels'. The top–levels and their children can be combined using defined relationships.

2.1.3.4 GTF Data Pool

With GTF, the structure of the numerous software applications and databases would be accessible in a homogeneous and compatible manner. A set of GTF Translators would provide a single access point to all models and data, see figure 2.

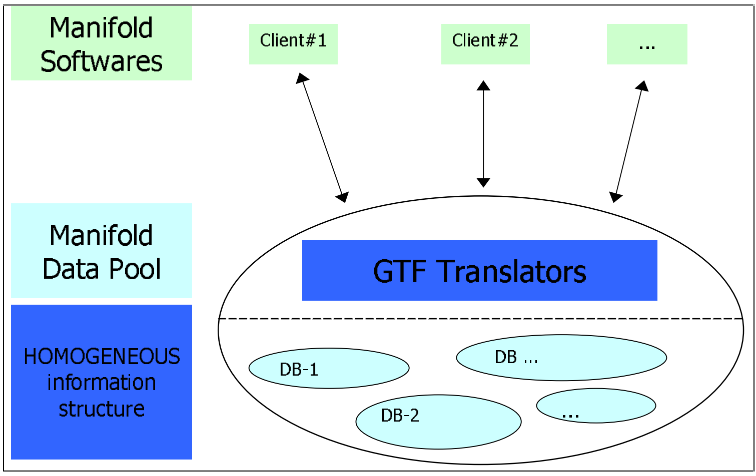


Figure 2. *GTF data pool.*

The numerous databases can either be restructured according to GTF's conceptual model. Or a specific GTF Translator for each database could be developed providing a homogeneous and single access possibility.

2.1.3.5 Implications / ramifications of GTF

The impact of GTF has many ensuing commercial and practical benefits:

23. Synergy effects emerge from the possibility of transferring knowledge between systems.
24. It will be possible to compare different models' results (and their quality) as the models can be used on the same data(base).
25. Model users may avoid to (re)create their own databases over and over again like in the past, but will have access to standard data(bases).
26. Data(bases) will gain in quality as time passes, because the data providers will have an incentive to update their databases regularly and properly, since only the 'good' databases will be used.
27. Users will request new models or combination of models, which previously could have been denied by the consultants, because of lack of transparency on the business.
28. The clients / users will have the possibility of choosing and combining models.
29. People dealing with problems appearing in different working areas can exchange information, e.g. to analyse side effects when changing from a higher to a lower aggregation level.

All these effects will have a vigorous impact on research in the modelling and other fields.

2.1.3.6 Comparison with other models and formats

GTF is designed to be a general conceptual model (for data) and a format, mainly addressing the demand for strategic - and hereby often multi-modal – transport models. GTF can describe domain specific objects, e.g. by using sub-classes. But this is not predefined as detailed as in some specialised formats, e.g. to describe details in rail switches or turn lane geometry at road intersections.

For comparison, figure 3 suggests the domain of application of GTF and other models and/or formats.

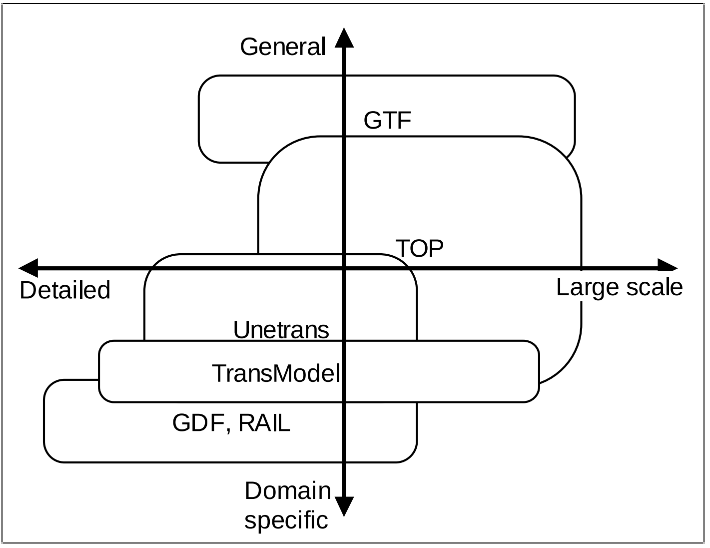


Figure 3. *Domain of different models and formats.*

The Transport Object Platform (TOP) is a conceptual model and its implementation for ArcGIS 8 systems. It has been inspired by the work in the BRIDGES project. TOP is in its predefined version less general than GTF. But its object model is completely open; users can add objects – even parent objects – that inherit, connect, or relates to other objects in the model. TOP is mainly developed as a general model for multi-modal transport. Furthermore it includes domain specific objects, e.g. turns at road networks, stops and terminals in transit networks, and complex demand for freight transport. Nevertheless it is less detailed than GTF in the way, that it only considers topological objects. It is up to the user to define attributes.

The UNETRANS transport data model (http://www.ncgia.ucsb.edu/vital/unetrans/) was also developed in relation to the ArcInfo GIS (and with some coordination with TOP). UNETRANS is, to a higher extent than TOP, a pre-defined model. Even many attributes are predefined. Its focus is mainly infrastructure (rather than transportation), with details especially concerning road networks. On the other hand the description of topological relationships in public transport is less comprehensive than in TOP.

The European TRANSMODEL for public transport (CEN-norm prENV 00278021) is a detailed – as well as large-scale – oriented model for public transport – especially bus-networks. It is more comprehensive within this domain than e.g. TOP version 1.0.

The European GDF-format mainly for road traffic (http://www.ertico.com/links/gdf/gdf.htm) has a higher degree of pre-definitions than UNETRANS. Although GDF is a format, and UNETRANS a model, their underlying conceptual models have similarities, as GDF was reviewed before defining UNETRANS

The RAIL model being developed as a counterpart to TOP is a detailed object-oriented model for rail infrastructure building on ArcInfo. It includes a number of domain-specific objects such as switches, signals, control sections, blocks, etc. RAIL is being co-ordinated with TOP as some TOP-objects can get aggregated information from RAIL, and RAIL can disaggregate information from TOP (e.g. the delay along a path from a rail-simulation can be aggregated into the TimeTable within TOP).

Finally, commercial modelling packages could be classified within the framework in figure 3. The main difference is that their formats are less open and that the models are predefined to a large extent. Detailed domain-specific software are, e.g. rail and road simulation, packages. Most transport modelling software is fairly general, with different specialisations from detailed modelling to large-scale. Some are even comprehensive covering multiple scales. GIS-packages are typical general packages, with few predefined domain-specific objects for the transport sector (Neither for detailed nor for aggregated purposes). This was the background for the development of TOP, RAIL and UNETRANS as extensions to ArcInfo. TransCAD (developed by the US firm Caliper) is an exception, since it both has GIS and modelling capabilities tailored for the transport sector (http://www.caliper.com/tcovu.htm). In addition, 'GIS-like' features are emerging in some commercial transport modelling packages.

2.1.4 ENTITIES IN GTF

This section introduces the fundamental classes that are the foundation of the GTF conceptual model. The transport data that is covered is primarily data used in strategic transport models. Thus it covers interurban, regional or international travel on all transport modes for both passengers and freight. More specifically, the meaning of 'Transportation Entity' in this paper is:

30. Transportation = 'The act of moving passengers or freight in space.'
31. Transportation Entity = 'All that produces (generates or attracts), enables or hinders movement of passengers or freight.'
32. Transportation Relationship = 'The connection between two transportation entities.'
33. Transportation Attribute = 'A quality or feature of a transportation entity that is a central part of its nature distinguishing its instances.'

The definition of e.g. TransportProduction in GTF contains not only the raw data, but also the meta-data, e.g. 'statistical source = EUROSTAT, type = statistics'.

The definitions above cannot be used for direct implementation. The goal of these definitions is to be able to define a conceptual model of transportation and not to implement this data model. The implementation is left to eventual providers who have to adopt GTF as one of the exchange formats of their software/model.

A Terminator is a virtual point for input & output (source & sink) of movement in networks. It is associated with Zone, which contains the TransportProduction of an area. In many transport models, the concept of a centroid is used to describe the same as Terminator in GTF. However, since centroids in some GIS implicitly are the geometrical centre of the zone (rather than the activity based centre like in transport model), the more general word Terminator is used.

A Terminator is connected to an infrastructure network through a ConnectorLink. The ConnectorLink is the virtual description of the impedance(s) that is needed in average to enter / leave a Zone and thus creating inter-zonal transportation / movement called LinkFlow.

A LinkFlow is the result of TransportProductions generating and attracting movement across the limits of Zones. It can therefore be described as a connection (relationship) between two Zones. This is 'flow' e.g. in the sense of demand for transportation. Thus, a LinkFlow is a connection between Terminators with information about the amount and types of flows (vehicles etc.) between the two Terminators. 'Flow' in the sense of observed movement is an attribute in the GTF Conceptual Model attached to a Link (or Segment).

A Node performs two functions in transportation modelling. The first function is to relate (connect) a Zone to some point in the network as access and egress points for mobility (the Node being a Terminator). The other function is that of being a Junction in the transport network. For generalisation purposes, Zone is a derived class from Node, too, as Zone's can be starting / ending points for Links not only Terminators.

A Link is a topological relation between two Nodes. The Nodes in turn usually are associated to specific geographical co-ordinates in real world space. But this is mostly needed for visualisation and presentation purposes.

Following this kind of logical decomposition and analysis, 10 top-level classes were as mentioned defined in the GTF: Node, Link, Mode, Vessel, Chain, DynamicSegmentation, Alternative, Unit, Group and Meta. Using these high-level definitions further child-classes are defined in the conceptual model. The next table gives definitions for each top level.

Class Name	Description
Node	The generalisation of the concept 'start or ending point of Links' and thus a generalisation (class) of the Terminator, Junction and Zone classes. Exactly two Node classes determine the generic class Link. This secures a more homogenous view on the problem domain.
Link	The Link class is not only an abstraction for all types of infrastructure network links, but it incorporates the connections to Zones (through their Terminators). Terminators, Junctions and Zones in different combinations act as Nodes to define three possible types of Links: 1. The Segment (LinkInfrastructure) connects two Junctions in the transport network 2. The LinkConnector between a Junction and a Terminator describes the disutility to reach (any) point in the Zone from the main transport network 3. The LinkFlow between two Terminators or Zones is a Link that holds the flow information that results when two Zones to describe the

Class Name	Description
	movement between two areas in space. For technical reasons this class is actually named 'MatrixElement'.
Mode	A Mode is the type of immobile infrastructure used by Vessels for the transportation of Units from Zone to Zone or between Junctions etc. i.e. on Links.
Vessel	Vessel is the abstraction of everything that moves on Links. In transportation models typical Vessels are cars, trains, aeroplanes, trucks etc.
Chain	The Chain represents the abstract concept of sequence of Links or activities. For example, a child class is Service that provides a traveller with the means to travel with relevant choices already made in advance by the service operator. The Service class defines the type of service, the used carrier Vessel(s), the level of security attributed to this type of service, and the timetable for the service.
Dynamic-Segmen-tation	Contains information of milestones, e.g. their position (distance from starting Node and distance form ending Node) and other data that is attached to a specific point on a Link.
Alternative	Transportation models use choice alternatives (e.g. usage of road or rail or air mode for transportation etc.) to describe the situation the behavioural units face in certain situations. From a transportation modelling point-of-view the networks (groupings of Nodes, Links etc. which form a logical whole) need often to be distinguished according to different 'main modes'. To broaden the definition, the more precise term Alternative defines 'choice alternatives'.
Unit	Units define the type of element being moved or transported (The purpose of the movement or the date / time schedule of a movement are stored in Meta.)
Group	The class can be used to group any class, class instance in order to define "result sets". This class is not like the others in the Toplevel. It is simply for grouping purposes. To add a level of semantics for the grouping one of the children classes should be used.
Meta	Metas are objects to define meta-information describing dimensions of measurements etc. The Metas can be used to associate dimension information with all/any other class instance.

Figure 5 depicts the top-level objects and their relationships in an UML diagram.

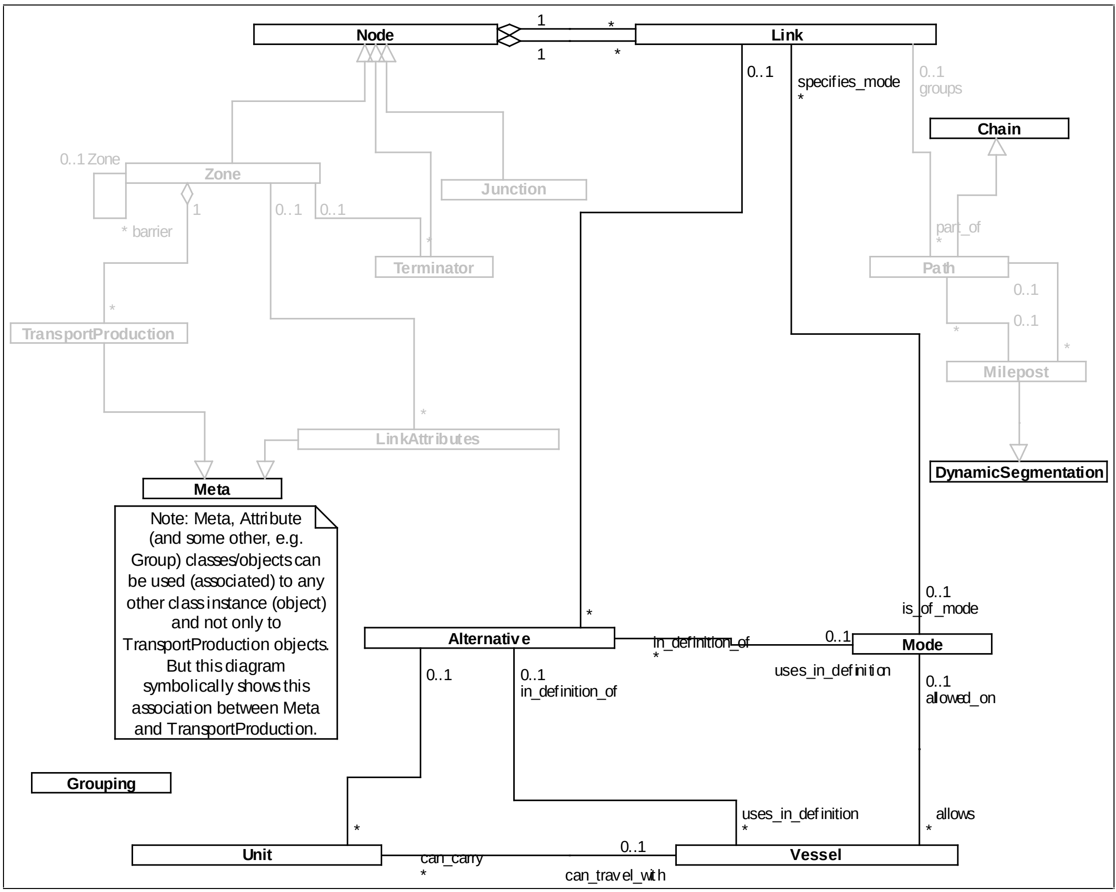


Figure 5. *Overview of the GTF model class and relationship structure.*

Note that all classes described in the model are instances of the class GTFObject that has a GIS pointer and a KIF conceptual pointer. The GIS pointer should be used to point to (an external) GIS object, e.g. a contour of polylines object. The KIF (Knowledge Interchange Format) can be used to contain a piece of text in KIF syntax. This can be used to describe some knowledge information according to KIF, e.g. 'f(origin, destination)=time + cost + weather' or knowledge of an abstract nature like 'for travelling business people time is much more important than cost'. This kind of information can be described formally in KIF, why it is understandable and can be processed by computers.

2.1.5 Using the GTF-format

From the description of the requirements of systems supporting GTF follows that modelling-data needs to be transferred across different platforms, mainly Windows and UNIX platforms. The structural system requirements are depicted in figure 6 (Mandel & Ruffert, 1999).

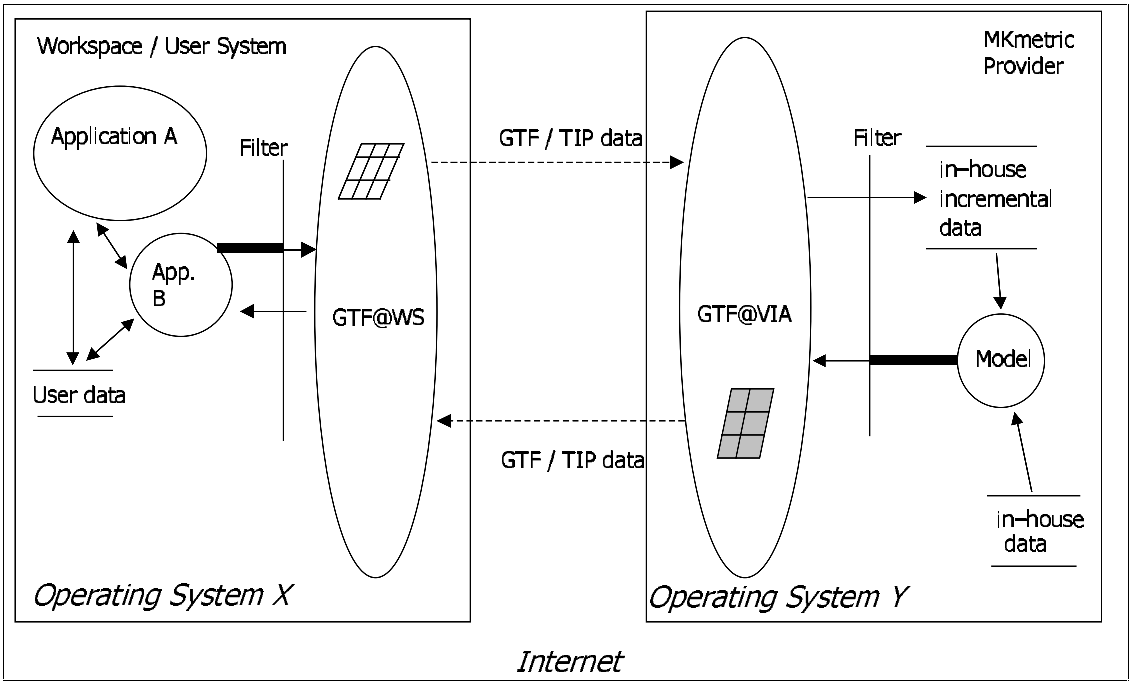


Figure 6. *Typical Exchange Structure.*

2.1.5.1 Workflow when exchanging information

As an example of a typical workflow, a user modifies his local ‘User data’ through his system. The user then formulates a request for a model, and the data to be used. A filter is used to make sure that only relevant data (-data not unknown to the model provider) gets translated by the GTF@VIA Translator (e.g. VIA = MKmetric’s model package). The resulting GTF file is transferred to the user’s account at the model provider’s server. There the data from the GTF file is extracted and incorporated as incremental data into the data already available at the provider’s (in–house data) site. The complete data is then fed into the chosen model according to the TIP information in the GTF file and the requested computations are done. The requested results are extracted by the filter and translated into GTF by the GTF@VIA Translator. The user’s system gets notified that the requested results are ready for download at the provider’s site. The user downloads the data. The user can then view the results with his favourite applications.

The consequences for the actual structure of a GTF file are:

34. Cross-platform / human-readability: A non–binary code must be used. The choice has fallen to the ASCII code, because this format has the least problems when being exchanged between heterogeneous platforms. ASCII also

has the additional effect that a GTF file in ASCII can also be read and understood by a human, e.g. in case of problems.

35. Segmented & Self-describing: As the data and control information to a model needs to be put together by the user's system the exchange format must be very flexible and powerful.

2.1.5.2 Transportation–data Interchange Protocol (TIP)

GTF specification also includes a number of commands that can be issued to a model provider's GTF enabled system. These are part of the GTF file and will enable a model provider to process the GTF data file so that the requested answers are computed. This is necessary, because the GTF conceptual model alone does not contain any information on what shall be done with the data. TIP is a generalisation of 'usual' commands (queries) to a transportation model. The development of the first version of TIP is based on the classic four-step transportation model. The commands can independent of the actual model or the model's philosophy be issued to the model in order to produce intermediate data or final model results. These results can then be passed through a filter defined in a TIP command file that is part of a GTF-XML file. The filter extracts data from model results corresponding to the user's query, and notifies the user's system that the requested results are available for download from the model provider.

2.1.6 THE TRANSPORT OBJECT PLATFORM – TOP

The Transport Object Platform, TOP, is an extension to the ArcInfo GIS. TOP has mainly been developed for the domain of public transport, but can be extended to other domains as well. TOP includes an object model (UML), developed within the object library in ArcInfo (typically inheriting features from ArcInfo objects, although TOP also include new parent classes). Data can be assessed through the objects of TOP (software application) or through the ArcInfo user interface, while the actual choice of database is flexible (ArcInfo support a wide variety of commercial databases). Compared to GTF, TOP, also includes a number of methods for the maintenance of the model, e.g. for editing, updating and visualising data.

2.1.6.1 Reasons for developing TOP

Public transport systems rely not only on a given infrastructure; it is also dependent on the available rolling stock and the possible timetable. Despite the interdependence between these elements, public transport companies often structure their data in a non-holistic way; e.g. by making separate departments responsible for infrastructure, timetable and rolling stock data, respectively. This tendency is strengthened by the deregulation of the public transport sector in many countries, making different companies responsible for data of the same transport system.

For this reason, data are often placed on different software platforms; Timetable and rolling stock data in different relational databases, infrastructure data are often divided in tabular data stored in a relational database, and geographical data stored GIS (Nielsen *et al.*, 1998). Some data are even stored in closed proprietary formats inside transport modelling software packages.

The distribution of data across multiple platforms makes it difficult for planners to construct models that fully utilise the available data because of inconsistencies between the different data platforms and conceptual models. This encourages *ad hoc* approaches to the tasks of translating and loading data into the models.

Furthermore, most data models are non-intelligent, in the sense that they do not prevent the existence of inconsistent data. The lack of proper visualisation and editing tools also contributes to the data inconsistency, since complex features - e.g. transfer links at terminals - are not treated explicitly as unique objects.

Many of these problems are similar to the reasoning behind the development of GTF. Other relates to the task of building and editing models of public transport.

2.1.6.2 Proposed Solution

With the introduction of object-oriented GIS based on standard relational data-bases, an elegant solution to these problems is now possible. The answer is to create an intelligent, rule-based, open and extensible object-oriented model.

Making a model intelligent and rule-based involves building functions (methods) into the model itself, rather than into the client of the model, e.g. into a transport modelling package. Based on defined rules, these functions can ensure data integrity at all times.

Making the model open (and non-proprietary) makes it generally accessible. Establishing a general model independent of existing transport model software can make the model serve as the intermediate step between raw data and data in the transport model.

Building the model based on object-oriented GIS and standard relational database technology makes it possible to use state-of-the-art of-the-shelf tools for editing, analysis and visualisation, including visualisation of non-physical – but geographical linked - objects, such as turns, transfers at public transport terminals and timetable data.

Overall, this new approach makes the highly time consuming data related steps in transport modelling easier and thus more cost efficient. In addition, consistency is enforced by the built-in functions in the objects. This greatly improves data quality and eases quality control.

2.1.6.3 Objectives

The objectives with TOP were twofold:

To be a functional GIS-based model for public transport. This included a conceptual model, development of a corresponding data model with objects for each type of topologic element, and finally implemented methods related to the objects to make the model functional, e.g. editing and updating methods, visualisation routines, query functions, and user interface.

Hereby to demonstrate that GIS-based object-oriented approaches are feasible today to model complex transport systems. This may launch new initiatives concerning other domains, e.g. rail infrastructure models, freight networks and terminals, and air systems.

As such TOP is a platform to be used for transport planning and modelling. At present, TOP is used to maintain the Copenhagen Ringsted model's data foundation (Nielsen *et al*, 2000). However, ongoing work extends TOP with transport related methods, e.g. assignment models. Being a practical tool – although a very general one – TOP is more than a data exchange format or data model, since methods are built into the objects in TOP.

2.1.6.4 The Conceptual Model of TOP

In the following, the main conceptual model behind TOP is described (figure 7). The full conceptual model – with 34 elements - is described in Nielsen & Frederiksen (2001a). The conceptual model reflects the preliminary design process and is the basis of *UML diagrams* used to describe the actual software objects in TOP (see Nielsen *et al*, 2001c). These consist of separate diagrams describing inheritance, relationships, connection rules and object functions.

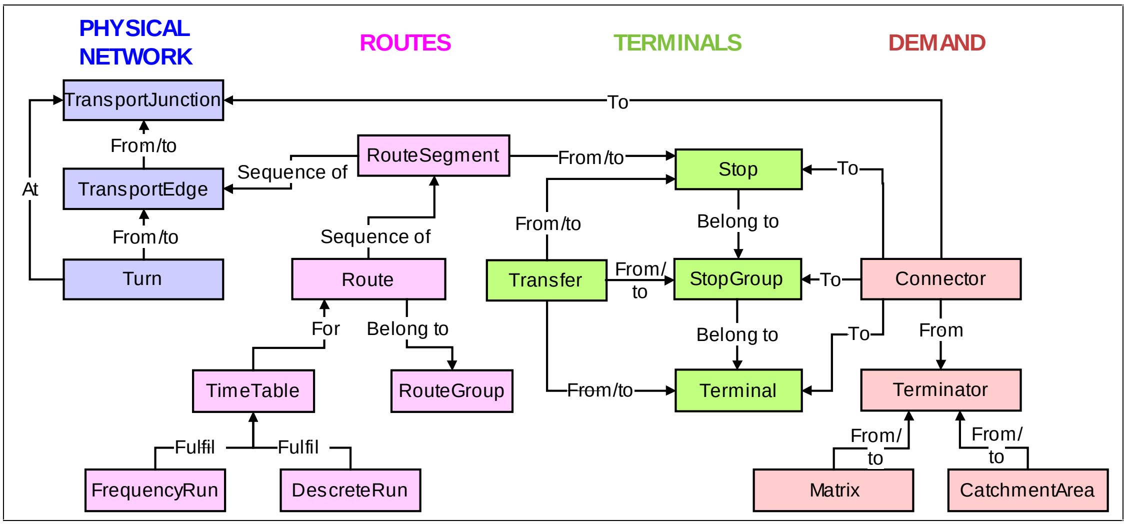


***Figure 7** Conceptual overview of the TOP object model*

In the following, object class names are written in Italics concatenated with capital letters starting the individual words, e.g. *TransportEdge*. Overall, the TOP consist of 4 main parts:

The Physical Network consisting of *TransportEdges*, *TransportJunctions* and *Turns*. Turns are mainly used to describe road networks. But they can also describe restrictions in e.g. rail switches.

- The Route Network describes scheduled routes on top of the underlying physical edges. A *Route* connects a series of *Stops*. A *TimePattern* shows which of the Stops along the route that are actually stopped at, and how long time it takes to reach the *Stop*. The run describes one specific departure. *Routes* can be grouped in order to describe a single public service with variations in the *Route*, *TimePattern* and *Runs*.

- Transit Terminals describe junctions in the public route network, and the possibilities of movement (*Transfers*) between stops within the terminal. *StopGroups* are aggregations (unions) of *Stops*, and *Terminals* unions of *Stops* and *StopGroups*.
- The Demand group of objects describe data elements commonly used in transport modelling. *CatchmentAreas* (e.g. zones) are used to divide a model area into a collection of aggregated elements. A *Terminator* is the network representation of the *CatchmentArea* in the form of a node. This is connected to the relevant *TransportJunctions* and Stops using *Connectors*. *Matrices* are used to store relevant information described on a *Catchment*-to-*Catchment* level, for instance number of travellers, travel time etc.

As part of the process developing the conceptual model of TOP, a review was made of the most widely used model applications on the market. This led to the addition of specialised objects to describe Terminals and Demand.

2.1.6.5 Comparison of entities in TOP and GTF

As mentioned, TOP and GTF have different purposes, why some of the objects are defined or named differently.

In TOP, Junctions and Links are named *TransportJunctions* and *TransportEdges*, because they have the more general GIS-objects *Junctions* and *Edges* as parents (which as well could be parent objects in facility networks, sewer lines, etc.).

In TOP, traffic can be produced (generated or attracted) not only by zones, but zones, lines or points all being sub-types to a *CatchmentArea*. As an example a freight model may describe factories and storage facilities as points - not zones (e.g. some European industry sectors have very few producers and stores, as car factories, and television producers). Also transportation of dangerous goods may be analysed at a point level, e.g. nuclear waste going from power plants, to a treatment facilities, and back to final storage. In GTF the class “Group” serves this purpose of assembling information (objects from different classes). The sub-types (derived classes) add a level of semantics. Like this, there is a sub-class “Catchment” which adds this semantic information to the grouping.

Furthermore, demand has its own class of objects (refer to Nielsen & Frederiksen, 2001a), including *ComplexDemand* for trip chains visiting several *Terminators*, e.g. for the use in activity based passenger models, freight models, or logistics problems. This can be mapped to the GTF class “Chain” and its sub-classes. The matrix concept from TOP was incorporated into GTF by renaming and changing the focus of the “Link Flow” class, which now is the “GTFMatrixElement” class which in turn is aggregated into the “GTFMatrix” class.

Concerning *Terminals*, the main difference is, that a *Stop* may not need to be a *TransportJunction* in the network. A bus stop may e.g. have been digitised independently, and has accordingly other coordinates than the road centreline (in reality, this is also the case since the pole is not standing in the middle of the road). In this case, the bus stop can be ‘linked’ to the nearest location at the road centreline by a reference, without the road centreline being split into two *TransportEdges* with a *TransportJunction* between.

A *Stop* may be connected to several networks / *TransportEdges*, e.g. a bus-stop to a road-centreline, a pedestrian path and a *Connector*, or an airport to a *TransportJunction* in the airline network, as well as to *TransportEdges* in the Rail and Road networks.

The concept of Terminals etc. basically can be mapped in GTF by using the super/sub association that every GTF class automatically inherits, since every GTF class is of type “GTFObject” which contains this association. This association allows one GTFObject to be part of another GTFObject or be associated to other GTFObjects that make up the parent object. Like this a hierarchy or a network of associations between GTFObjects can be created.

2.1.7 SUMMARY, DISCUSSION AND CONCLUSIONS

GTF is an acronym for ‘Generalised Transportation-data Format’; with the goal of standardising the information used by transportation modelling software for the purpose of electronic data interchange (EDI). The GTF specification uses already defined standards wherever possible in order to maximise acceptance and to minimise redundant work. To accomplish this the GTF specification comprises the following parts:

A standardised definition of transport information, but without constraining the possible information to any specific domain. This is called the *GTF Conceptual Model (GTF-CM)*.

A standardised set of commands to run models and to retrieve relevant data. This is called *Transportation-data Interchange Protocol (TIP)*.

A standard format for arranging data in a file used for *Electronic Data Exchange (EDI)* and a standard protocol for exchanging the data file. For this XML is used. (GTF-XML)

This paper addressed primarily the main components of the GTF conceptual model (section 4). As the technical descriptions of the other two components, and details on the conceptual model are comprehensive, they are only briefly described in the paper.

During the discussions within the EU-project BRIDGES, followed by the thematic network SPOTLIGHTS, it was realised, that formulating a fixed data model was virtually impossible at the European level, due to the large differences in conceptual models, data definitions and software solutions found in different countries, within different domains (e.g. transport sectors), and at different levels of aggregation. Realising this, it was decided to implement a flexible format, which can be read and interpreted from any software platform (given GTF-translators have been implemented).

2.1.7.1 GTF specification

The GTF specification was developed to enable model providers to offer their transportation models' results in a standard fashion. Subsequently, this enables computer systems to present the results in the form a user wishes. A complete system furthermore should assist the user with the tasks of finding appropriate data and appropriate model providers to answer a user's transportation query.

The specification does not cover everything in detail, but tests showed that models of urban transport, freight and passenger models, special models for shipping, road specific information on load or damages, schedules as well as indicators or indexes can be handled by GTF.

2.1.7.2 Technical development of GTF

During the work with GTF – and discussions with model providers, model users and modellers – it has been realised, that the balance between flexibility and predefinitions a format is difficult.

Without offering the possibility to add sub-classes or new parent classes, one risk that GTF cannot contain the richness of a certain model, whereby it becomes useless for certain data sets.

However, if many users add their own extensions to GTF it become less general with the risk of being a set of tailor-made formats for which all other modellers need to develop specific versions of their GTF-translators. The ultimate problem with this may be different GTF definitions of models that in fact are conceptually equal. Hereby, GTF would de facto degenerate into several – related – exchange formats.

The solution is not easy. However, the best approach seems to:

36. Extend GTF with new core-objects if several models need these.
37. Extend GTF with parent and sub-classes 'labelled GTF-versions', when more than one model need additions that describe the same conceptual phenomena.
38. Extend GTF with tailor-made additions only for phenomena that are contained in one model only. These additions should only be sub-classes, since other models that do not use this richness can interpretation an exported data-model using the implicitly given parent classes.

At the last SPOTLIGHTS meeting, it was e.g. decided to include several of the objects from TOP as 'labelled GTF-versions', which may at a later stage be 'promoted' to core-objects (e.g. the widely used linear references and turns).

A procedure of submitting 'change requests' to http://gtf.mkm.de or spotlights@mkm.de is in the process of being installed and formalised.

2.1.7.3 Future use of GTF

In our point of view, there is a widespread waste of resources within the modelling community due to inconsistent data and lack of reuse of existing data. However, modelling is a complicated field. And the present version of GTF became very complicated in order to capture the complexity of transport models. Even as such, it covers mainly more well-known model types.

On the other hand, resources for modelling are often low. Furthermore, some software products trap their customers by using closed proprietary data models, and/or insufficient exchange routines. As such, the vendors have neither economic nor businesslike reasons for implementing a unifying exchange format (in this context the OpenGIS consortia by the leading GIS-vendors is a revolutionary step within the GIS-community).

There are also business-like and political reasons that hinder the exchange of modelling data, e.g. between competitive rail operators, or certain regions that do not want other organisations to question modelling results they use to advocate for certain subsidy from the government or EU.

Besides technical issues within GTF, these organisational and political issues have to be solved, before the visions with GTF can migrate into practice.

2.1.8 REFERENCES

Booch, G., Christerson, M., Fuchs, M. and Koistinen, J. (1999) *UML for XML Schema Mapping Specification.* http://www.rational.com/media/uml/resources/media/uml_xmlschema33.doc

Brown, David (1997) *An Introduction to Object-Oriented Analysis: Objects in Plane Language*, John Wiley & Sons

Budd, Timothy (1997), *An Introduction to Object-Oriented Programming*, Addison-Wesley

EUROSTAT (1996) *GESMES 93 – Exchange of Multidimensional Statistical Arrays and Time–series Data*, Volume 1: Guidance to Users, Volume 2: Reference Guide, EUROSTAT, Luxembourg.

Mandel B. & Ruffert E. (1999) *GTF Final Report.* MKmetric GmbH. EU-project BRIDGES.

Mandel B., Ruffert E. (2000) Generalised Transportation-data Format (GTF): Data, Model and Machine Interaction; paper presented at the *1st ITEM Workshop; Montreal/Canada; 10/2000.*

Mandel B. & Ruffert E. (2001) *GTF Specification.* MKmetric GmbH. EU-project SPOTLIGHTS.

Marchal, Benoit (1998) *XML by Example*, Que; ISBN: 0789722429

Nielsen, O. A., Israelsen, T. & Nielsen, E. R. (1998) *BRIDGES TO GIS – Methodology,* Deliverable D5 & D6. BRIDGES Contract No PL96-1138. EU, DG7, 4th Framework Programme.

Nielsen, O. A., Hansen, C. O. & Daly, A. (2000) A Large-scale model system for the Copenhagen-Ringsted railway project, **Proceedings** *of the 9th International Conference on Travel Behaviour Research***,** vol. **12**, Application Workshop 4: Large scale model systems. Gold Cost, Queensland, Australia, July. (To be published in a revised version in

Travel behaviour Research: The Leading Edge. Book edited by David Hensher. Pergamon press. Chapter 36, pp 597-616. 2002).

Nielsen, O.A. & Frederiksen, R.D (2001a) Rule-based object-oriented modelling of public transport systems, **Proceedings** *of the 9th World Conference on Transport Research*, July, Soul, Korea.

Nielsen, O. A & Frederiksen, R. D. (2001b). GIS-based object-oriented modelling of transport networks, **Proceedings** *of the European Transport Conference (PTRC)*, Proceedings, Seminar on Methodological Innovations, September, Cambridge.

Nielsen, O.A., Brun, B., Frederiksen, R.D., Grevy, B., Israelsen, T. Poulsen, M. & Skriver, J. (2001c) Data Modelling for Transportation Infrastructure Objects, **Proceedings** *of the Twenty-first Annual ESRI (Environmental Systems Research Institute) International User Conference***,** San Diego, 2001.

Rumbaugh J, Blaha M, Premerlani W, Eddy F & Lorensen W (1991) *Object–Oriented Modelling and Design*, Prentice Hall, New Jersey.

UML, resource (documentation of UML, XML as well as example of the use of XML) http://www.rational.com/uml/index.jtmpl

2.1.8.1 Projects

BRIDGES, 'Building Bridges between Digital Transport Databases, GIS Applications and Transport Models to Develop ETIS Software Structure' (contract no. ST-96-AM-1138), on behalf of the Commission of the European Community – DG VII, 1997-1999

SPOTLIGHTS(TN); 'Scientific forum for making advanced transport models fully transparent to end-users, open and more integrated into policy-making'; on behalf of the Commission of the European Communities– DG-Energy and Transport; Actual Cost Contract No.: 1999-TN.10941; 2000-2003.

2.1.8.2 Acronyms and definitions

Conceptual model	The description of objects and their relationships in a model, i.e. the structure of a model – not its implementation
Data Base	The data in a specific model stored electronically
Data Format	Specific format for exchanging data
Data Model	A conceptual model, with precise definition of all objects, their data definitions as well as each data-element
EDI	Electronic Data Interchange
EC	European Commission
EU	European Union
EUROSTAT	The statistical bureau of EU
GTF	Generalised Transportation-data Format
GTF-CM	GTF-Conceptual Model
GTF-XML	GTF's XML-based exchange format
ITS	Intelligent Transport Systems
KIF	Knowledge Interchange Format

TIP	Transportation-data Interchange Protocol
Model	The implementation of a data model in a specific software system including all needed data (and implicitly build in methods as well)
TOP	Transportation Object Platform
UML	Unified Modelling Language
XML	eXtensible Mark-up Language (Metagrammar for interorganizational communication around the Internet)

2.2 World Conference on Transportation Research 2001

See the proceedings of the world conference on transportation research, Korea 2001. Presentation by Prof Otto A. Nielsen.

2.3 UNETRANS: UNIFIED NETWORK-TRANSPORTATION DATA MODEL

See the proceedings of the Turin consultation http://www.ncgia.ucsb.edu/vital/unetrans/. Presentation by Prof Otto A. Nielsen.

2.4 GTF Copenhagen internal meeting / workshop

Participants: MKmetric, DTU/CTT, WS Atkins, ScanRail.

Goal of spotlightsTN/GTF

- Get the modelling community in Europe (in the World) to "establish the necessity of a common exchange format"
  - ⇒ first suggestion of an operational format

- contact modellers / software providers
- discuss GTF Goal
  - ⇒ means for discussion: GTF-CM
  - ⇒ discuss GTF-CM:
    - Email spotlights@mkm.de
    - Forum gtf.mkm.de
    - Mailinglist (in approx. 2 weeks)

Outcome: Agreement & MoU

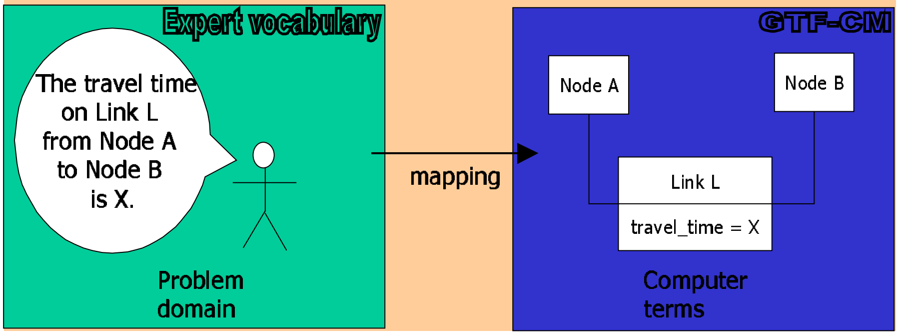
15th August 2001 Copenhagen MKmetric

GTF definition

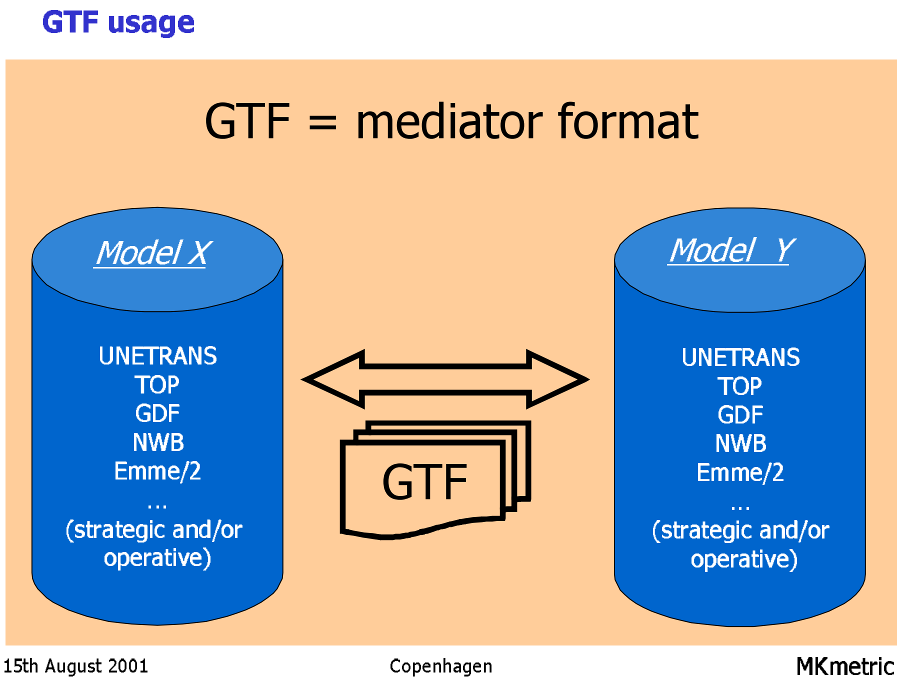
GTF = for the exchange of strategic transportation modelling information

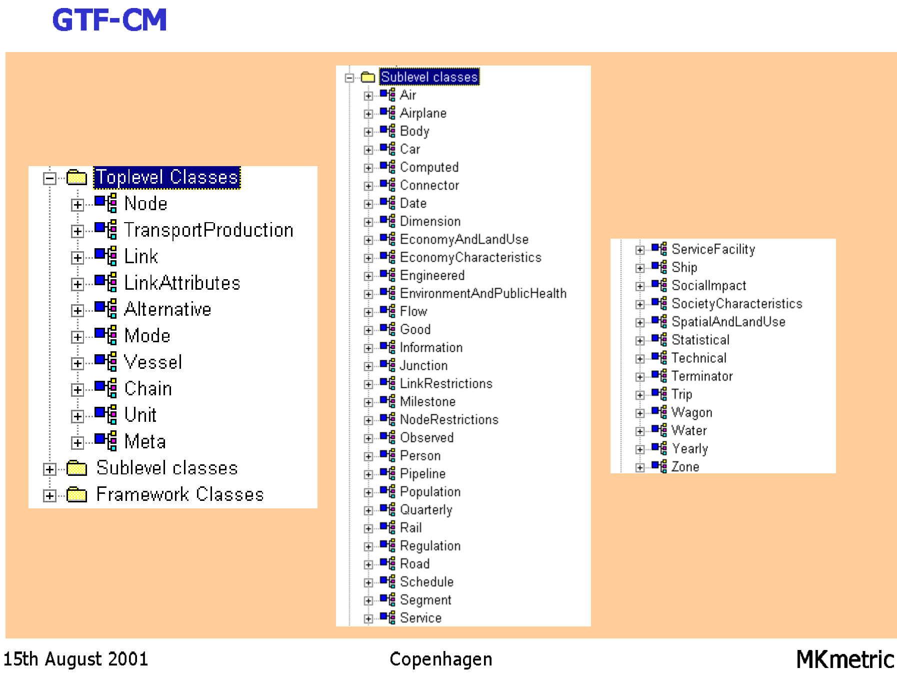
⇒ means: GTF-CM

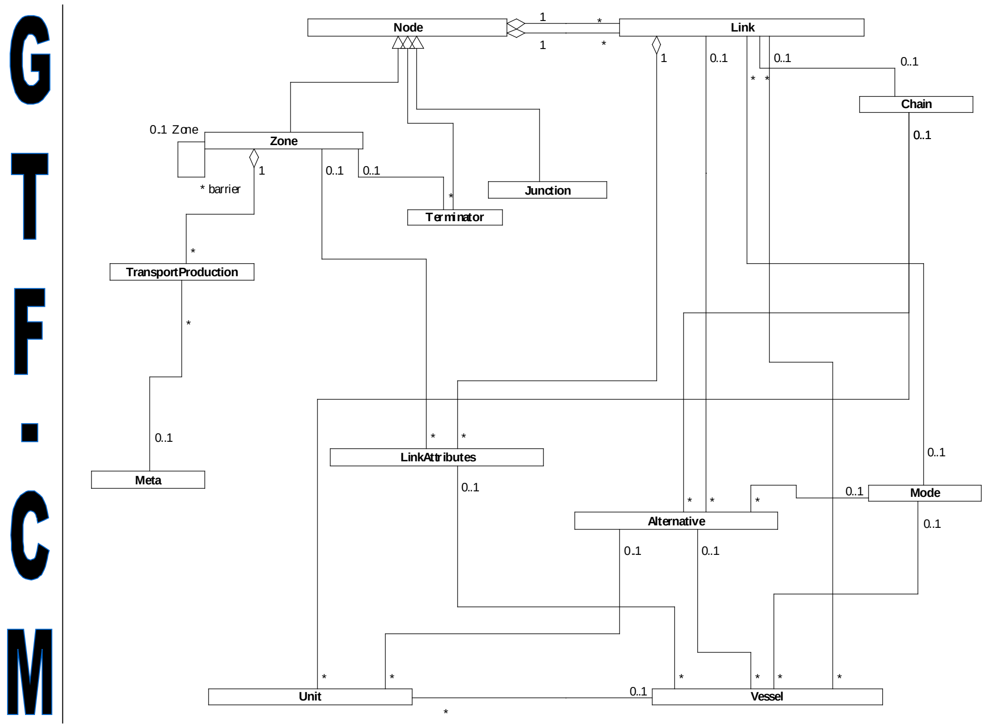
= capture the vocabulary of the problem domain in computer terms



15th August 2001 Copenhagen MKmetric


GTF usage
GTF = mediator format
Model X
UNETRANS
TOP
GDF
NWB
Emme/2
...
(strategic and/or operative)
GTF
Model Y
UNETRANS
TOP
GDF
NWB
Emme/2
...
(strategic and/or operative)
15th August 2001
Copenhagen
MKmetric


GTF-CM
Toplevel Classes
Node
TransportProduction
Link
LinkAttributes
Alternative
Mode
Vessel
Chain
Unit
Meta
Sublevel classes
Framework Classes
Sublevel classes
Air
Airplane
Body
Car
Computed
Connector
Date
Dimension
EconomyAndLandUse
EconomyCharacteristics
Engineered
EnvironmentAndPublicHealth
Flow
Good
Information
Junction
LinkRestrictions
Milestone
NodeRestrictions
Observed
Person
Pipeline
Population
Quarterly
Rail
Regulation
Road
Schedule
Segment
Service
ServiceFacility
Ship
SocialImpact
SocietyCharacteristics
SpatialAndLandUse
Statistical
Technical
Terminator
Trip
Wagon
Water
Yearly
Zone
15th August 2001
Copenhagen
MKmetric



Principles (main)

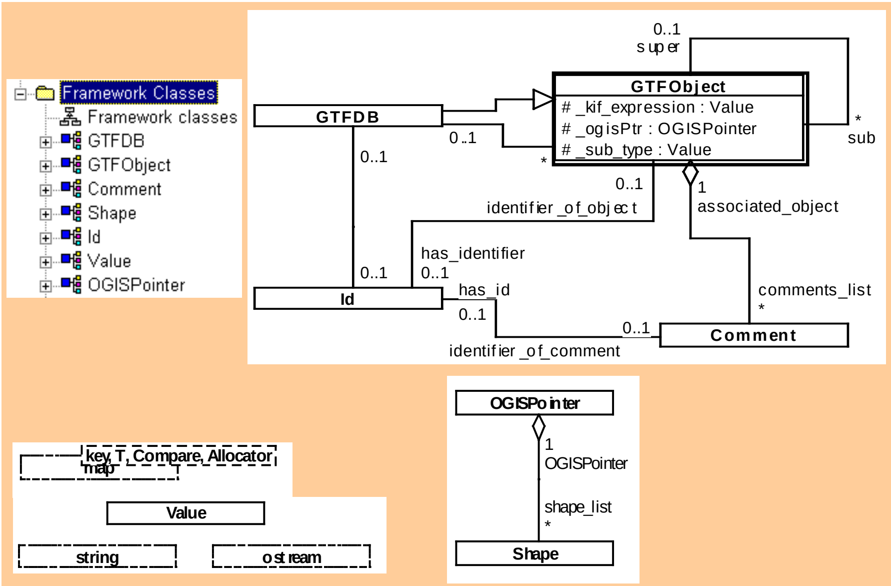
- all information is on Link
- each object: tree structure (super/sub association)
- all problem domain classes are founded on a few framework classes
- the number of top level problem domain classes shall be small

Note:

the GTF-CM specification contains both problem domain classes (Toplevels + Sublevels) & Framework classes

15th August 2001 Copenhagen **MKmetric**

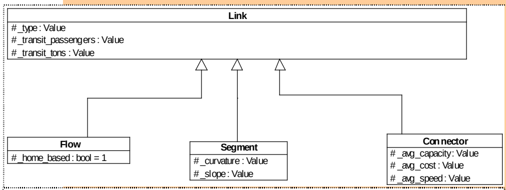
Principles / Framework classes for GTF-CM specification



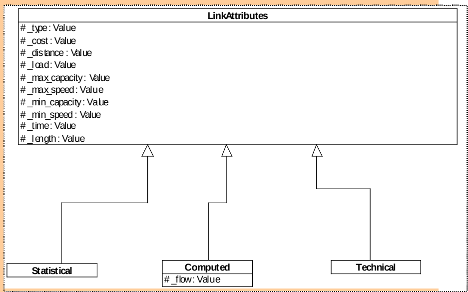
15th August 2001 Copenhagen MKmetric

Design decisions (example)

Link / LinkAttributes



Types of Links



Characteristics of Links

Design problem:

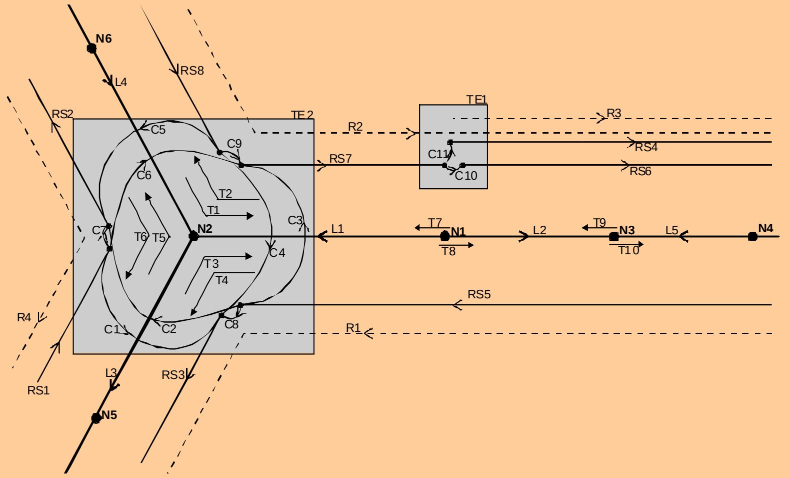
To fit both hierarchies of information would imply to use "multiple inheritance"

Decision:

This would lead to problems trying to map the GTF-CM onto relational databases, so just associate them

15th August 2001 Copenhagen MKmetric

Typical application

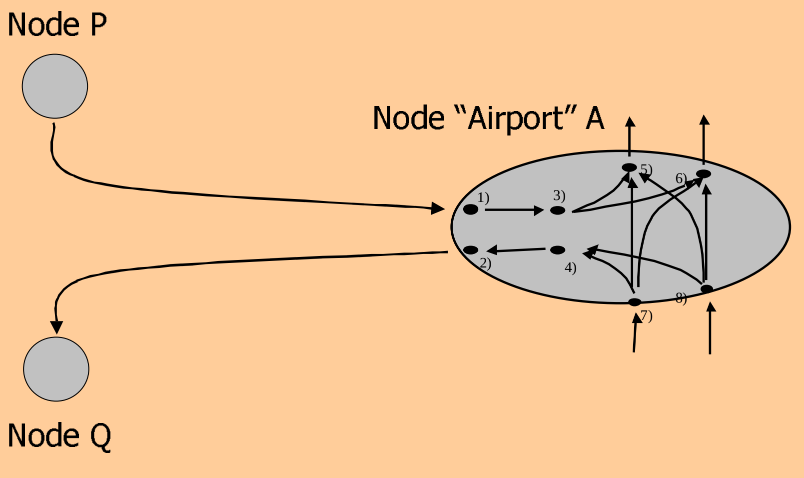


schematic view of a public transport terminal*

* taken from the BRIDGES/D4 report "Handling Traffic Modelling Networks in GIS" by DTU/TetraPlan

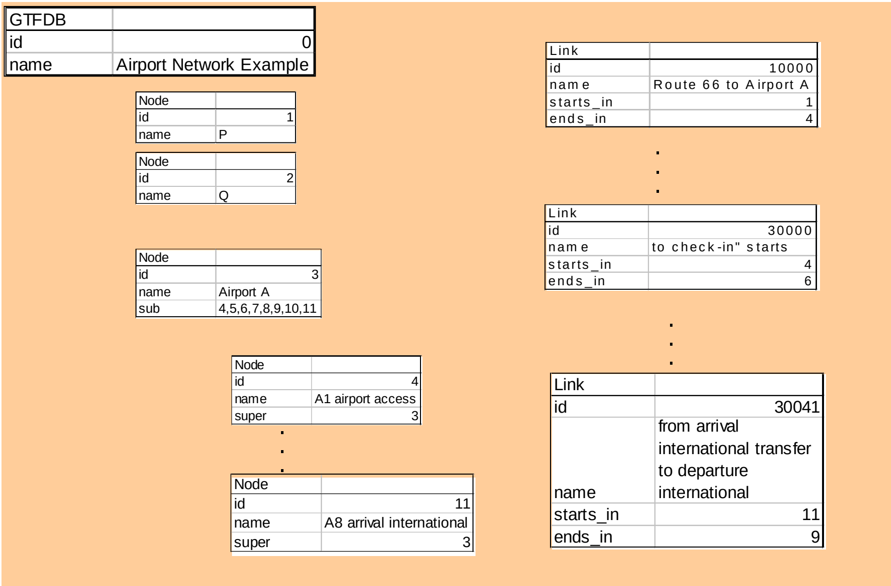
15th August 2001 Copenhagen MKmetric

Example application



15th August 2001 Copenhagen MKmetric

Example application: GTF objects



15th August 2001 Copenhagen MKmetric

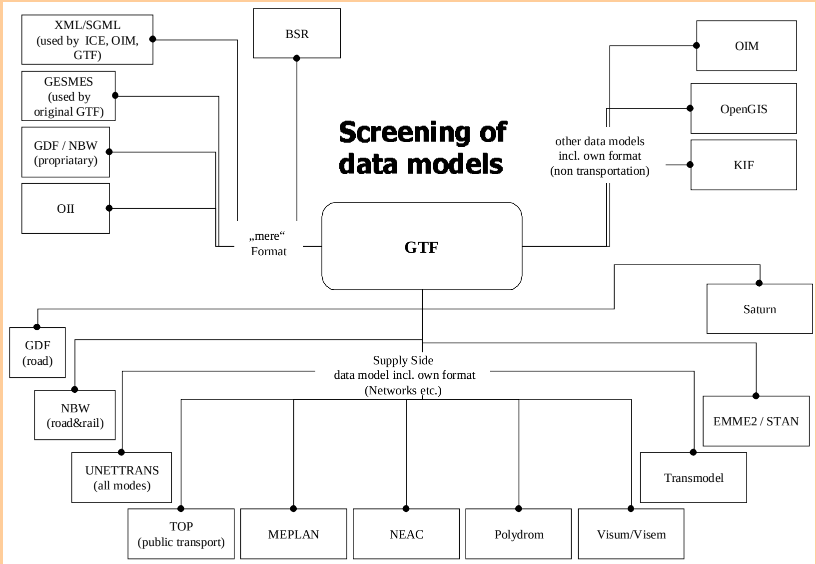
Example application: XML file

```
<!-- definition of the Airport Node-Network example-->
<GTFDB id=1 name="Airport Network Example">
        <N id="1" name="P">
        </N>
        <N id="2" name="Q">
        </N>
<!-- definition of the internal Nodes -->
<N id="3" name="Airport A">
        <N id="4" name="A1 airport access">
        </N>
        <N id="5" name="A2 airport egress">
        </N>
        <N id="6" name="A3 check-in counter">
        </N>
        <N id="7" name="A4 check-out counter">
        </N>
        <N id="8" name="A5 departure national">
        </N>
        <N id="9" name="A6 departure international">
        </N>
        <N id="10" name="A7 arrival national">
        </N>
        <N id="11" name="A8 arrival international">
        </N>
</N>

<!-- definition of the link from Node O to Airport A -->
<L id="10000" name="Route 66 to Airport A" starts_in="1" ends_in="4">
</L>
<!-- definition of the link from Airport A to Node D -->
<L id=20000 name="Highway 928" starts_in="5" ends_in="2">
</L>
<!-- definition of the internal links of Node Airport A-->
<L id=30000 name="to check-in" starts_in="4" ends_in="6">
</L>
<L id=30001 name="from check-out " starts_in="7" ends_in="5">
</L>
<L id=30010 name="to departure national" starts_in="6" ends_in="8">
</L>
<L id=30011 name="to departure international" starts_in="6" ends_in="9">
</L>
<L id=30020 name="from arrival national" starts_in="10" ends_in="7">
</L>
<L id=30021 name="from arrival international" starts_in="11" ends_in="7">
</L>
<L id=30030 name="from arrival national transfer to departure national" starts_in="10" ends_in="8">
</L>
<L id=30031 name="from arrival national transfer to departure international" starts_in="10" ends_in="9">
</L>
<L id=30040 name="from arrival international transfer to departure national" starts_in="11" ends_in="8">
</L>
<L id=30041 name="from arrival international transfer to departure international" starts_in="11" ends_in="9">
</L>
<GTFDB>
```

15th August 2001 Copenhagen MKmetric

What was done?



15th August 2001 Copenhagen MKmetric

What was done?

- email, personal phone calls to modellers
- email, personal phone calls to software providers
- update GTF Specification
- Web forum (eConference)
- Maillinglist

15th August 2001 Copenhagen MKmetric

Opinions from modellers / software providers (in general)

- would be nice to have GTF, but nobody forces us to do so
- we are quite content with the formats currently available
- too ambitious
- if I need to exchange some data I can do that on an one to one ad hoc basis, that's good enough
- proprietary formats protect our know-how and our market shares
- users urge us for more compatibility to any other product and other platform
- where is the market? what is it good for?
- no problem to write 1:1 translators and include them in our software if needed

15th August 2001 Copenhagen MKmetric

Opinions from modellers / software providers (specific)

- Different set of entities
  network, movements, stage, purpose, trip,
  journey / exit, daily activity program, vehicles,
  places, persons, institutions, observations
- Different paradigm
  interfaces & components
  description language about models
- Other focus
  "too complex" solutions are not useful because nobody understands them
- Criticism
  Specification document not good enough, too confusing
  Some names contradict well defined terms in trans. modelling

15th August 2001 Copenhagen MKmetric

Open questions

Main discussion points for Copenhagen meeting

- resolve confusing names

- dynamic segmentation
- complex demand
- catchment area
- corridors

- need “network” object?

- attributes?

15th August 2001 Copenhagen MKmetric

MINUTES OF COPENHAGEN INTERNAL GTF MEETING 15TH AND 16TH AUGUST

Day 1 at WS Atkins

1. Presentation of TOP
The concepts of TOP were presented and discussed in detail. Refer to [Nielsen & Rasmussen, WCTR-paper describing the conceptual model in TOP and Nielsen et.al. ESRI-paper describing its implementation]
The current implementation was presented as an application.

2. Presentation of GTF
The concepts of GTF were presented and discussed; the focus of the GTF Task and the GTF-CM (GTF-Conceptual Model) was discussed.

3. Comments / Opinions from other contacted Modellers / Software Providers were presented and discussed.

4. GTF-CM was discussed in detail and the concepts explained.

Day 2 at DTU

1. Proprietary structures <-> GTF-CM
It was discussed how TOP structures can be mapped onto GTF-CM structures and vice versa. The only difficult task was to embed public transport routes, especially Stops. A modification of this was agreed upon. Solutions on the other objects (complex demand, catchment areas, turns), where also found. These will be described in an updated version of the documentation.

2. Other proprietary structures
The Brussels meeting raised the question of how to embed a number of other proprietary structures.
It was found necessary to embed a new top-level class, “Path”, that can be used as a building block for a number of conceptual objects, e.g. the route between two matrix elements, public transport routes (TOP) and reference in Dynamic segmentation. A chain (e.g. TRIP-chain) is defined as a series of paths.

Results can be described as sub-classes of the new top-level class “Grouping”. Examples are Sheaf, Corridor (following Ulied's definition), Bundle, Spider, Scenario, Organisation.
It was agreed upon, that dynamic segmentation is seldom necessary for traffic modelling purposes at a level of aggregation larger than local. However, it was also agreed upon, that dynamic aggregation must be embedded none-the-less in GTF, since some data-sources store data in this way. A class for this was defined, using either LinkSegment or Path as the reference line. Dynamic segmentation can be of point type or defining a linear attribute between a from and to measure (milepost).

3. Attributes in GTF
The GTF-CM specification is to incorporate the concept of an "Attribute" class. The instances (objects) of this class can be used to attach attribute data to any other object in a GTF-Database. The GTF-CM will provide a list of default Attribute objects. It must be noted that these are to be preferred before any user defined Attribute is to be defined. The interpretation of these user defined Attributes will be described in an attached Comment.
In order not to make GTF to restrictive, the flexibility to add attributes was agreed upon.
The former top-level classes LinkAttributes and TransportProduction are hereby incorporated as subclasses to the more general Attribute class.

4. Generalisation of classes
A number of classes and relationships to other classes was generalised. As such, Attribute, Mode, Vessel, Unit & Meta can refer to any other classes in a more flexible way, e.g. both to Link, Node and Path.

5. GTF-XML
It was agreed that XML is the proper technology to define a concrete format for data of a GTF-Database (i.e. a database structured according
to a specific implementation of the GTF-CM).

6. Documentation

It was agreed upon to change the report (GTF-Specification) with an overview chapter, and chapter describing the main Top-Level Diagram, and chapters dedicated to each top-level class, with description of sublevel-classes as sub-sections within that chapter. With the number of main changes (as described above) as well as following changes in sub-classes (to comprehensive to describe in this note), the documentation will be rewritten significantly. Accordingly, it was agreed that MKmetric will write the next version, after which Otto Anker Nielsen will review and comment on it. When agreed, this new documentation will be sent for comment to the other partners.

7. ETC-paper
The presentation of GTF at the European Transport Conference was discussed. The structure of the presentation was outlined. It was agreed that the most recent changes should be presented, although the paper itself describes a prior version of GTF. The Barcelona meeting is mentioned, and the GTF-Website is referred to for newer versions of the documentation.

8. GTF Workshop
The upcoming GTF Workshop in October was discussed.
A list of potential invitees was agreed. These will be contacted, also directly by telephone, and will be part of the upcoming mailing list for a final intensive discussion about the GTF-CM (updated after this meeting, code name "Copenhagen version" v0.6r4).
This list will lead to a short list of invitees for the GTF workshop.

9. It was agreed that a “Memorandum of Understanding” should be suggested for the Workshop – and if possible signed during the Workshop. A draft of this MoU was agreed upon.

2.5 ITEM Workshop 1 – Montreal – 13th-14th of October 2000: Presentation Paper

ITEM WORKSHOP 1 – MONTREAL – 13TH-14TH OF OCTOBER 2000

Generalised Transportation-data Format (GTF)

— Data, Model and Machine Interaction —

Benedikt N. Mandel and Eduard Ruffert

MKmetric Gesellschaft für Systemplanung mbH,
Durlacher Allee 49,
76131 Karlsruhe, FRG
email: mandel@mkm.de and email: ruffert@mkm.de

Abstract

Exchanging data and information between (strategic transport) models and between models and other software, e.g. GIS, is always a very tedious, if even possible, task. There is always the problem of loss of information because the exchanged data only seemingly contains the information required and there is also always the problem of inhomogeneous and proprietary data formats forcing the users of the data to re-format and re-combine the data from scratch every time.

The solution to these problems is to understand that not only data needs to be transferred, but also the precise meaning of the data (meta-data). For the definition of information typically needed by strategic transport models a data model which precisely defines the data and information to be exchanged and for the exchange per se a standard format for electronic data interchange (EDI) is needed.

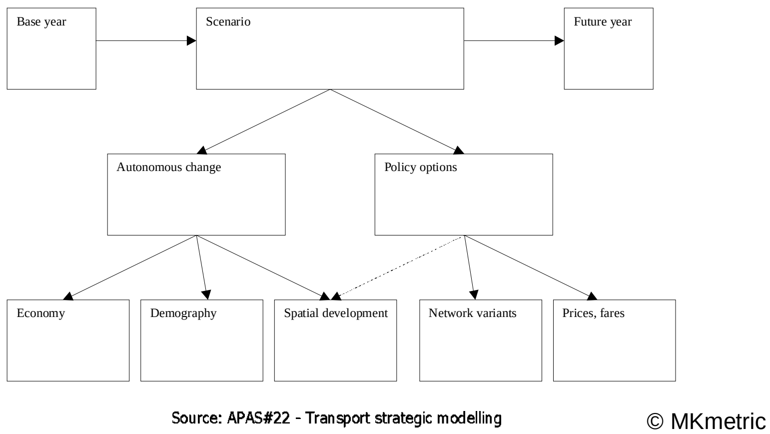
The data model must be an information model which enables a user (typically an applications programmer) to use the building blocks specified in the data model, to define precisely the data and the information contained in the data to transfer. The exchange format should be based on a standard. In addition, an interchange protocol should be defined in order to make the whole process of running a model and retrieving results automatic.

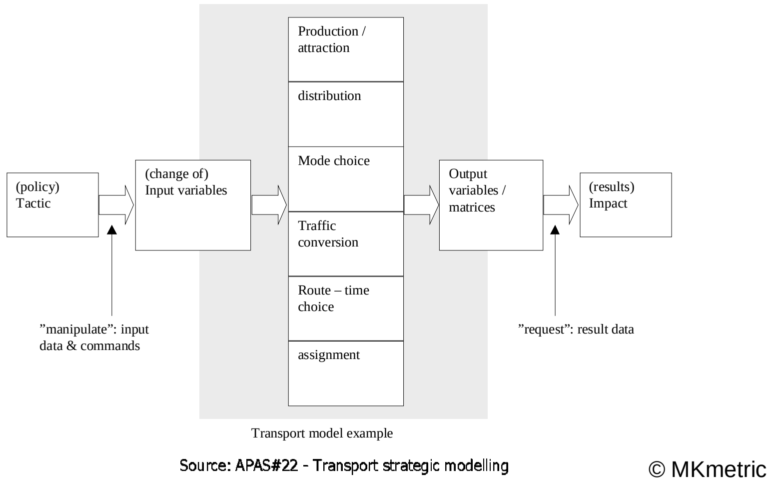
This paper defines such a data model (“Generalised Transportation-data Format” GTF) and proposes an exchange format (GTF-XML) based on standard XML which allows a software application (called a “GTF Translator”) to exchange information and data between models and between models and other software and proposes an interchange “language” to run models and retrieve results (TIP).

Keywords: exchange, interchange, format, protocol, EDI, data, information, XML, transportation, model, strategic

TABLE OF CONTENTS

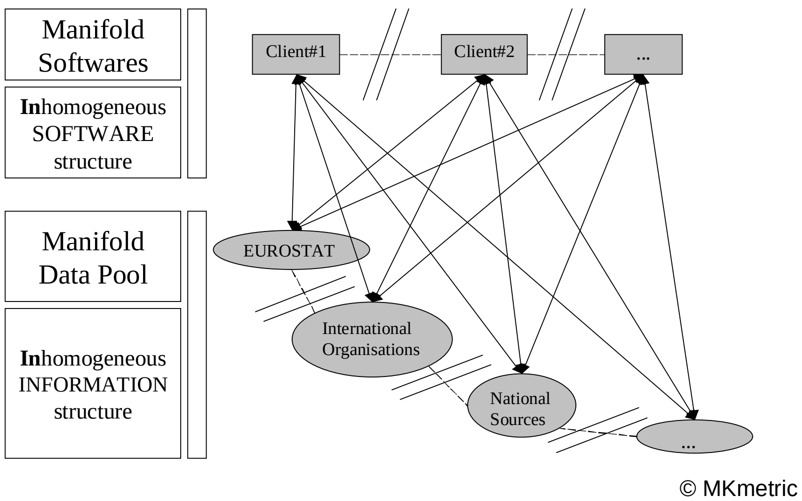
TABLE OF FIGURES





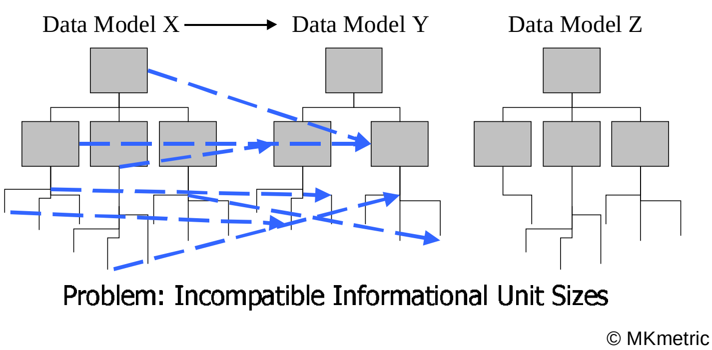
112

Fig. 17. Queries & Models 112



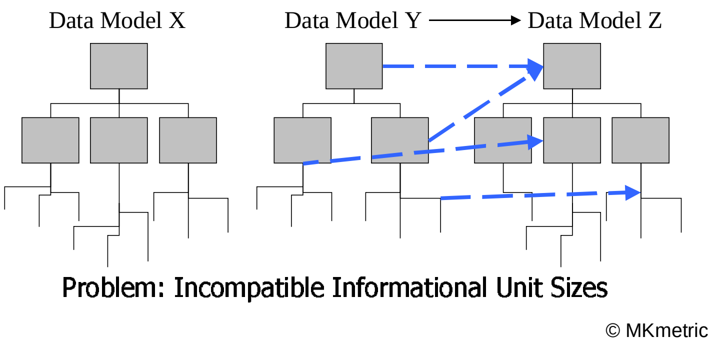
113

Fig. 18. Current Data Pool 113



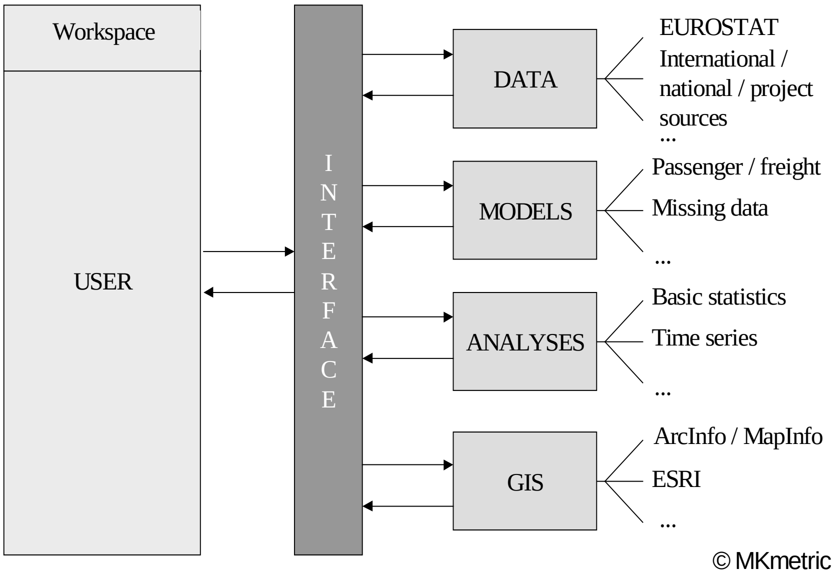
114

Fig. 19. Matching Data Models Step 1 114



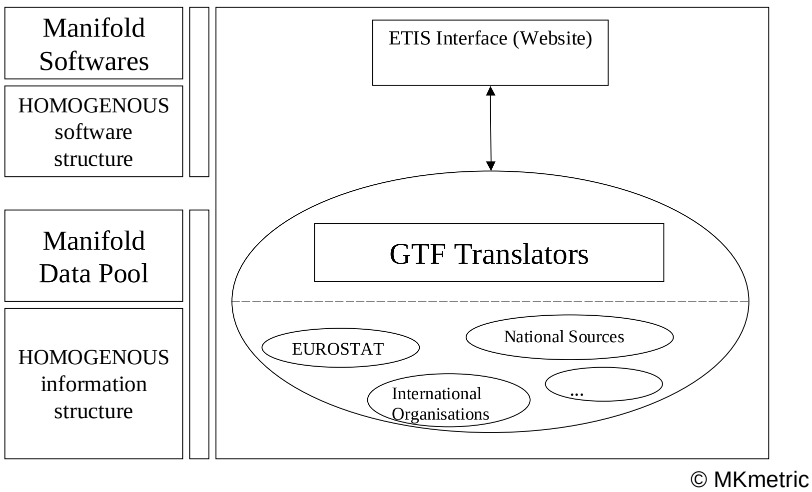
114

Fig. 20. Matching Data Models Step 2 114



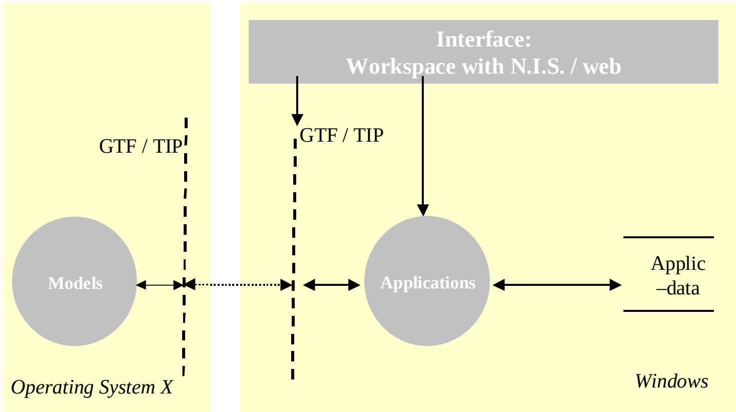
115

Fig. 21. ETIS Vision 115



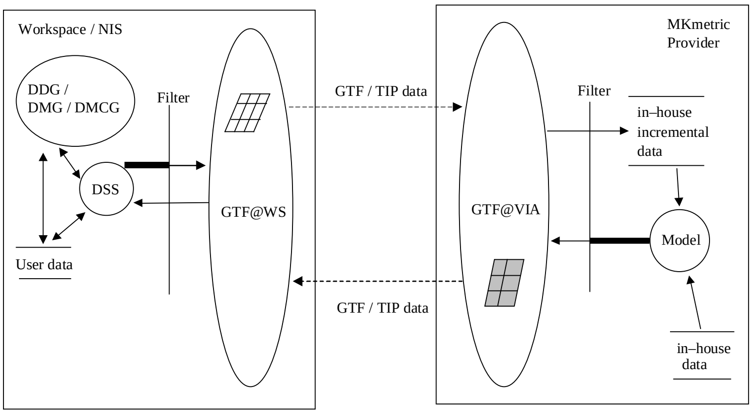
115

Fig. 22. ETIS / GTF Data Pool 115



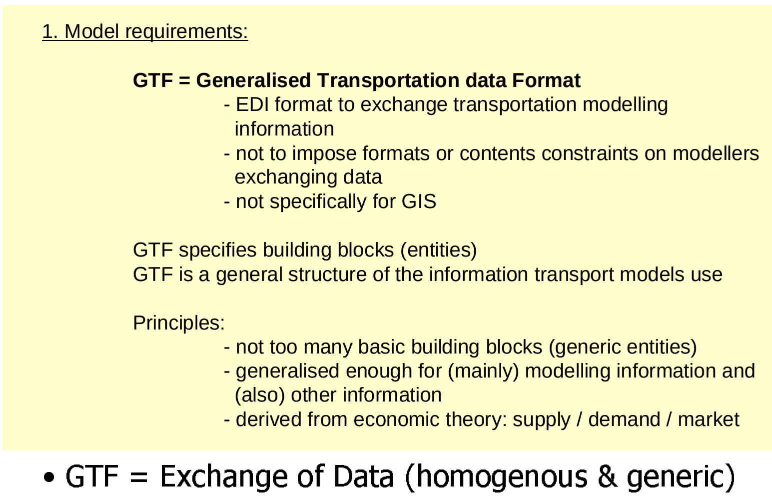
116

Fig. 23. ETIS Structure Overview 116



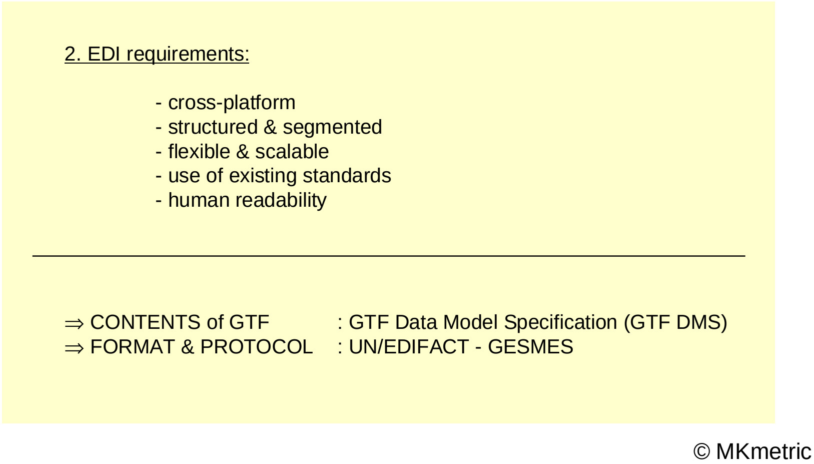
117

Fig. 24. ETIS Concept 117



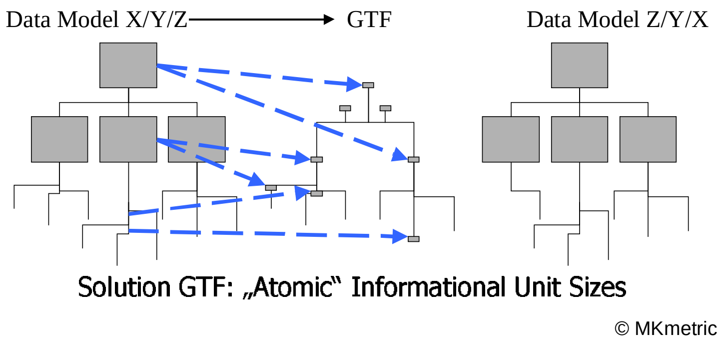
118

Fig. 25. Supporting Model Communication 1 118



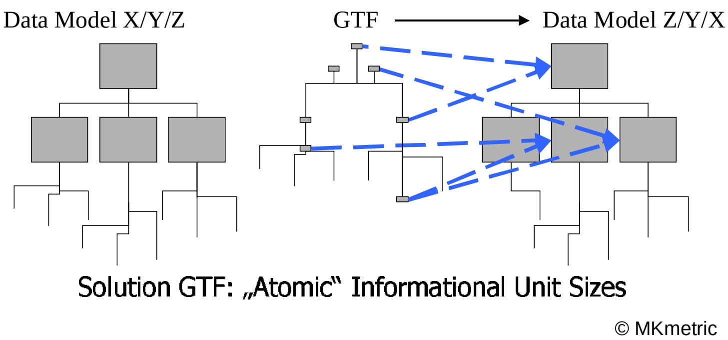
119

Fig. 26. Supporting Model Communication 2 *119*



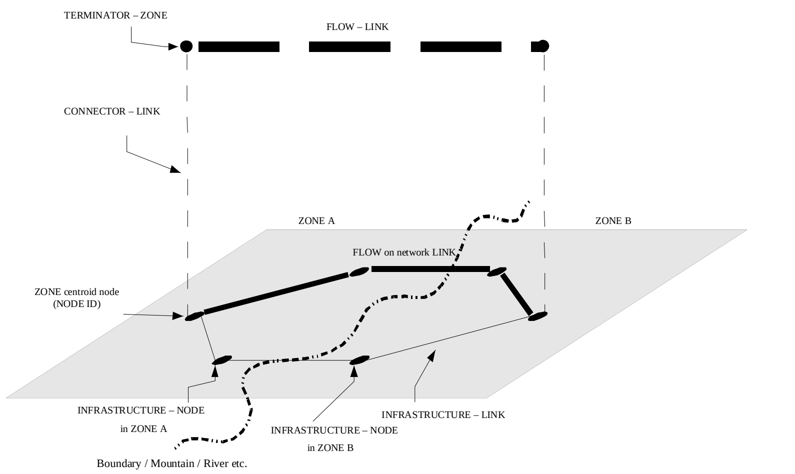
120

Fig. 27. Matching Data Models using GTF Step 1 *120*



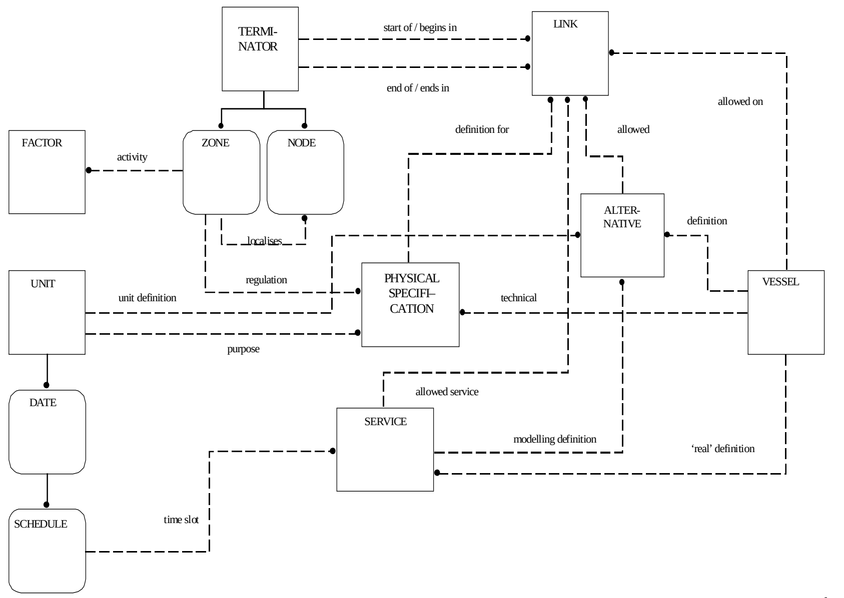
120

Fig. 28. Matching Data Models using GTF Step 2 *120*



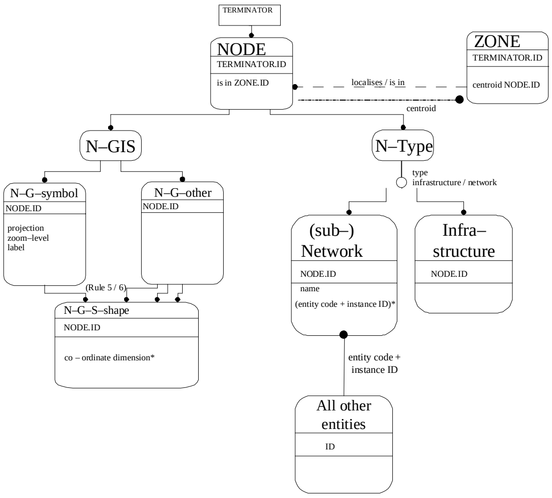
121

Fig. 29. GTF Data Model Overview 121



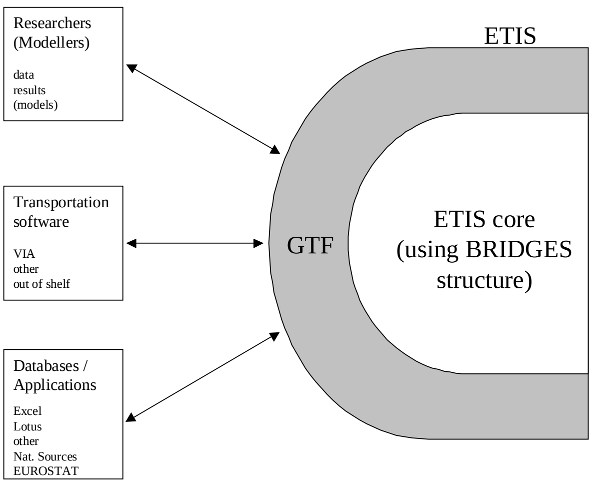
122

Fig. 30. GTF Data Model Entity-Relationship Overview 122



123

Fig. 31. GTF Example: Node 123



125

Fig. 32. GTF ETIS External Interface 125

• TIP = Transportation-data Interchange Protocol

= language

– to define queries to models

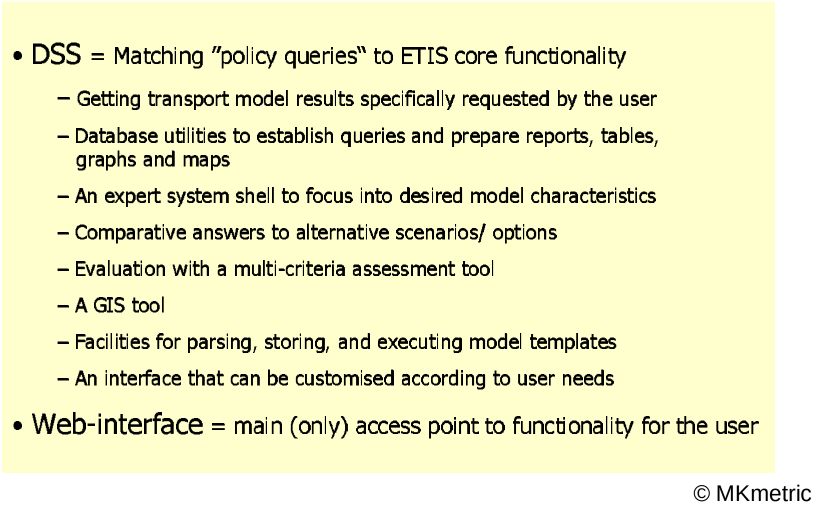
– to run models and

– to retrieve results

automatically

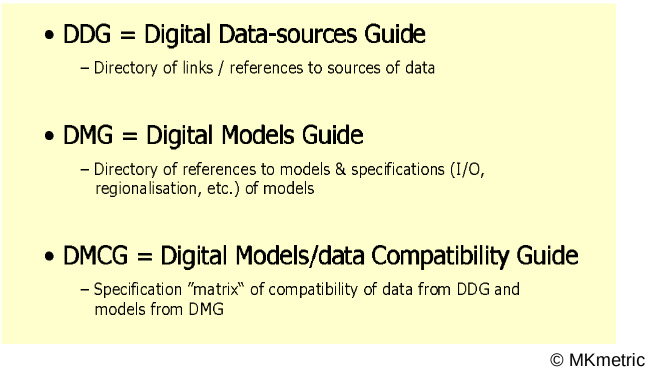
125

Fig. 33. Supporting Automation 125



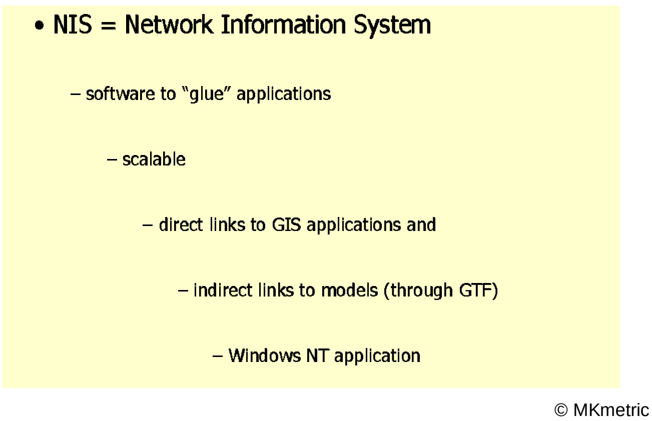
129

Fig. 34. Supporting Interface 129



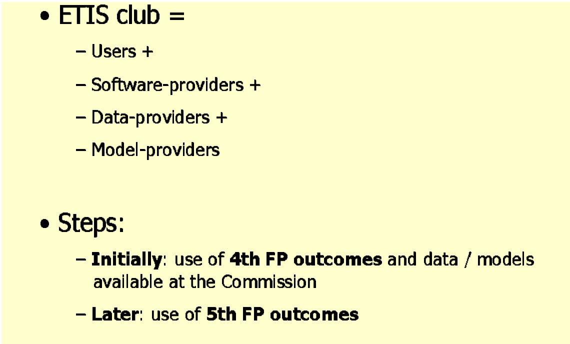
131

Fig. 35. Supporting Guides 131



132

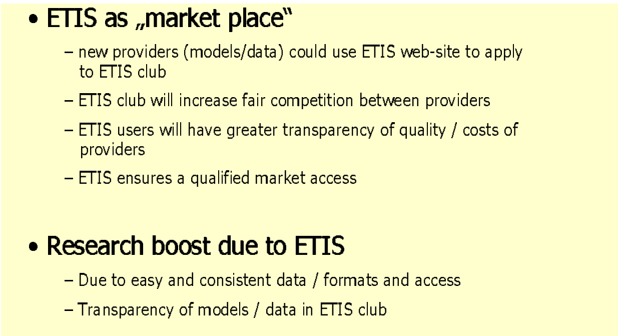
Fig. 36. Supporting Software Communication 132



© MKmetric

134

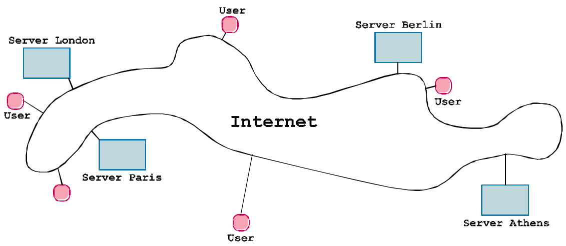
Fig. 37. Initial Implementation of ETIS *134*



© MKmetric

136

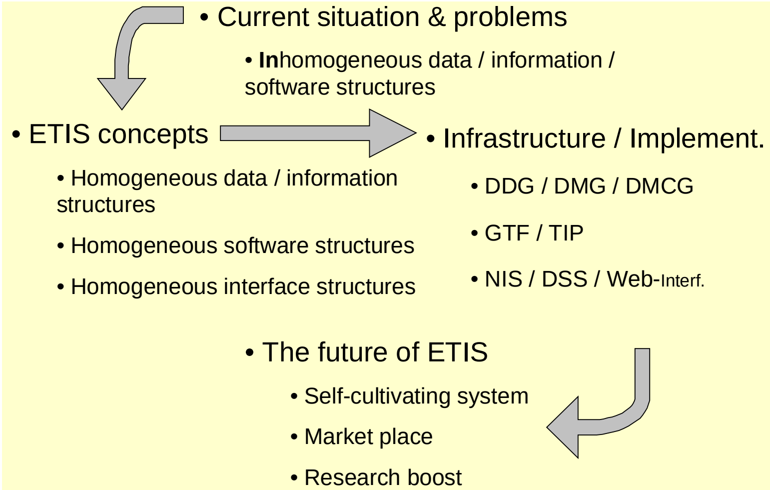
Fig. 38. Self-Cultivating ETIS Structures *136*



© MKmetric

139

Fig. 39. ETIS club Intra- / Internet Structure *139*



© MKmetric

140

Fig. 40. Summary *140*

2.5.1 Models, Data, Software and Policy Scenarios

Models are used to represent a view of a part of the real world in order to reduce the complexity of the interrelationships of the elements in the view and subsequently to make it simpler to derive conclusions and logical relationships between the elements in the view. Once satisfied with the model, i.e. the results from an input impetus are regarded to be sensible and realistic, these results are translated from the simplified view, back to the real-world part thus having results, e.g. a forecast, of the changes that would occur for an equivalent real-world impetus.

Very basically, a model is a formulation (e.g. mathematical and / or verbose – the mathematical formulation has the advantage of being able to be processed by computer) of the real-world view being analysed. “Things” that can be analysed could be for example, the passenger flow between a set of different cities (passenger transport model), the freight flow between a set of different production and sales points (freight transport model) or the price of stocks and bonds of a market, e.g. the DAX in Germany or the NASDAQ in the United States. These examples already show that models and their formulations – since they are used to describe the real-world – can be as diverse and complex as the real-world itself. Nowadays, the only limitation to the practical use of complex models is the sheer computational performance required when using complex models on very disaggregated data.

A very successful way and the current state-of-the-art of modelling parts of the real-world is by using mathematical formulations (and computer data structures) to describe

the elements and their relationships of the view of the real-world being analysed (– currently there are also a number of other concepts, e.g. using neural networks for modelling). And then to use aggregated or disaggregated data – depending on the specification of the model – to flesh-out the data structures and to run result procedures which in turn serve as input for the estimation of the model's parameters.
For a disaggregated model (e.g. passenger transport models) the basic input data required are socio-economic data of the elements (i.e. passengers), the sets of sources and sinks used (called zones, i.e. the areas of generation of transport demand, e.g. a zonal structure like the administrative zones of a country, which are areas populated by the elements) and structures representing the supply for transport (i.e. the infrastructures that support the ability for mobility, which are usually represented by network systems, e.g. a representation of road, rail etc. infrastructure). This information (the input data-set) is used to feed an estimation procedure of a model which in turn is used to describe the factors that influence the decisions taken by the elements. The output from the estimation are the parameters of the model which completes a model's practical specification – the mathematical formulation plus the estimated parameters based on the input data-set.
When one speaks of transportation modelling one is talking about two models that are used simultaneously to solve (or answer) a problem: a model of the transportation infrastructure (i.e. a model describing the supply side of transportation) and a behavioural model (i.e. a model describing the demand side of transportation). These models reduce the complexities of the real world into manageable chunks and, in principle, can be handled separately.
The infrastructure model defines networks (infrastructure supply side or abstract networks), vehicles (e.g. cars, trains, aeroplanes etc.), services (facilities for loading and unloading at a port) etc. that are based on the real world (observable) "things".
The behavioural model defines 1. abstractions of zones, zone features, choice alternatives etc., either in an aggregated or in a disaggregated fashion 2. the way that the formulated actors of the problem domain react and decide, given sets of choice options. This model is usually based on survey data. The more disaggregated the model (and therefore the required survey data), the more complex the model becomes mathematically. Disaggregated models have the advantage of being more accurate in forecasts and in their analysing behaviour. (At this level there are many connections to social science, because both try to explain differences in behaviour of groups based on their social, economic etc. differences.) The GTF data model [CHEN76] [NIST93] is

another reduction of complexity, making the modelling information manageable for EDI. The reduction is done by grouping and classifying the modelling information. For example, the concepts centroid node and intersection node are very different in the problem domain (and the usual models), but they share a common function of being ending points of links: centroids being ending points of flow-links and intersection nodes being ending points of infrastructure-links. These kinds of abstractions are the gist of data modelling and the contents of section “Generalised Transportation-data Format (GTF)”.

The current practice is to gather data (or databases) for models, to re-format the data, re-compose the information contained in the data and to apply the result as input to the model. This very often leads to problems because the information contained in the different databases used for the model aren’t compatible, meaning that after re-formatting and re-composition, the data set shows that vital information is missing. After the model is run and the outputs generated the practice is more and more to use a GIS (Geographical Information System) to visualise the results [BRIDGES]. This also almost always leads to problems, because the structure used by a typical GIS doesn’t contain information useable by a typical model. Also, there is the problem of matching the GIS structures and information to the model’s informational structure before being able to visualise the results.

This paper defines an information structure (which includes information for models and information for GIS) and an exchange format that ensures that information isn’t lost and that all the information contained in data is defined explicitly (– if the specification of the data model is followed accurately). The conclusions of this paper can primarily be applied to strategic passenger transport models but can be applied just as well to other models because all conclusions are derived from an abstract view of models, data, software used to implement models and software to view results, i.e. software to view the explained or forecasted flows between the modelled zones.

2.5.1.1 Current Situation & Problems

The usual use of a strategic model (e.g. for forecasting) is to define changes in the input data for each scenario to be analysed. The usual inputs that define “Policy Scenarios”, like economic, demographic and spatial developments as well as network changes and changes in prices and fares for the use of transport supply [APAS96] are depicted in Fig. 1.

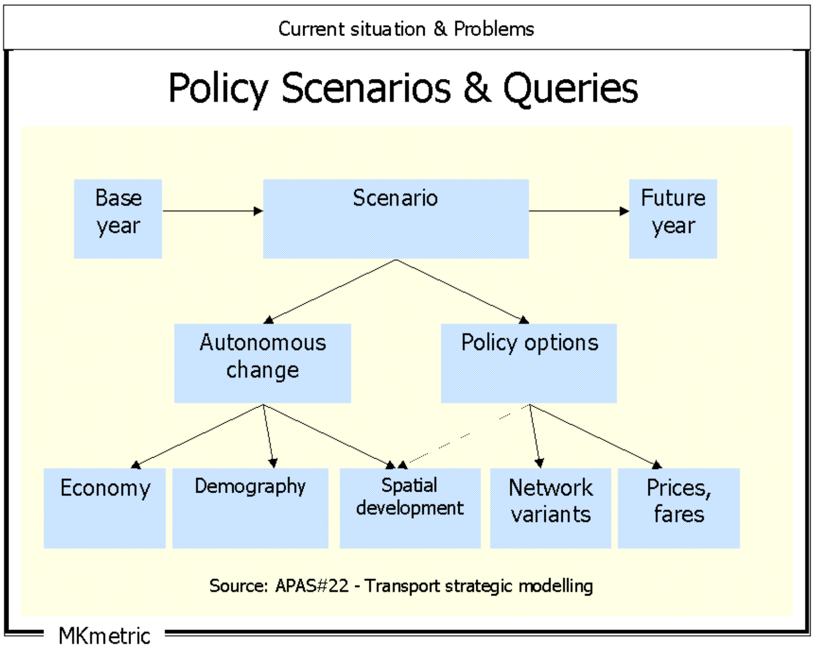


Fig. 1. Policy Scenarios & Queries

Fig. 1 shows that models are demanding on the amount of data needed. What also has to be ensured is the quality and informational contents of the data used for the model for estimation or calibration purposes. These two points are the main problems when trying to exchange input or output data for models. The exchange of data also includes acquiring data for models and transforming this data into the format and contents required by the model.

Another main aspect of using data or databases for different models is the informational contents of the data. Because, “data” can be defined as “the concrete value of a piece of information”. For example, a piece of data is “50” but the piece of information required by the model is “50 km/h”. The component describing the concrete value is a “fact”. The component describing the informational definition is the “meaning”. And only both together give a complete piece of information. Thus “information” can be defined as “fact” plus “meaning”. For the example, if one only has the fact “50” without the meaning “km/h” one would need to make assumptions about the meaning of the fact, which usually leads to errors. Because, “50” could also mean “miles/h” which would imply a different level for the data, as “miles/h” is about 1.6 times greater than “km/h”. This simple example shows that a model is extremely dependent on the correct information and not only the data.

Policy Queries & Models

The usual workflow, currently, is depicted in Fig. 2. The policy to be analysed using models is "translated" into model specific commands such that the concrete model, that is chosen by the user, is run and produces results – that are hopefully the results required for the user to make an objective decision for the problem at hand.
The figure shows that the definition of scenarios to be analysed by a model strongly depends on the manipulated information (and data) for the definition of a scenario. Using the example in the previous section, an increase of the speed for a type of an infrastructure (network) link (e.g. a road section) from "50" to "60" in "km/h" would be from "50" to "57" (approx.) in "miles/h". This problem cannot be overcome even if the scenario defines the increase as a percentage and not by an absolute value, e.g. increase by 10%.

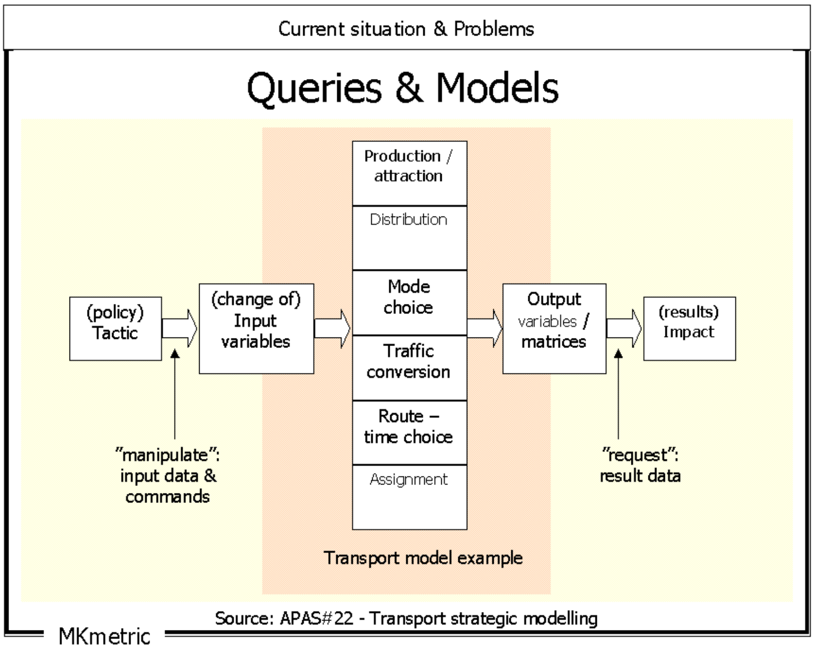


Fig. 2. Queries & Models

Current Data Pool

Currently, the structure of the numerous software applications and databases is inhomogeneous and largely incompatible with each other. Which, very frequently, leads to the problem (– rather the impossibility) of comparing results from scenarios based on different software applications and databases (see Fig. 3).

The problems mentioned in the previous sections also apply to the databases of the results from models (– not only on the input databases). Most models have specific database formats for the output. The output information is aggregated or composed according to the needs of the model package or the modeller using the model. The problems also apply, because it is often a wish to compare results from different models (from different projects) or to use the output from one model as input to another model. In these cases, the same problems of re-formatting and re-composition of the data arise.

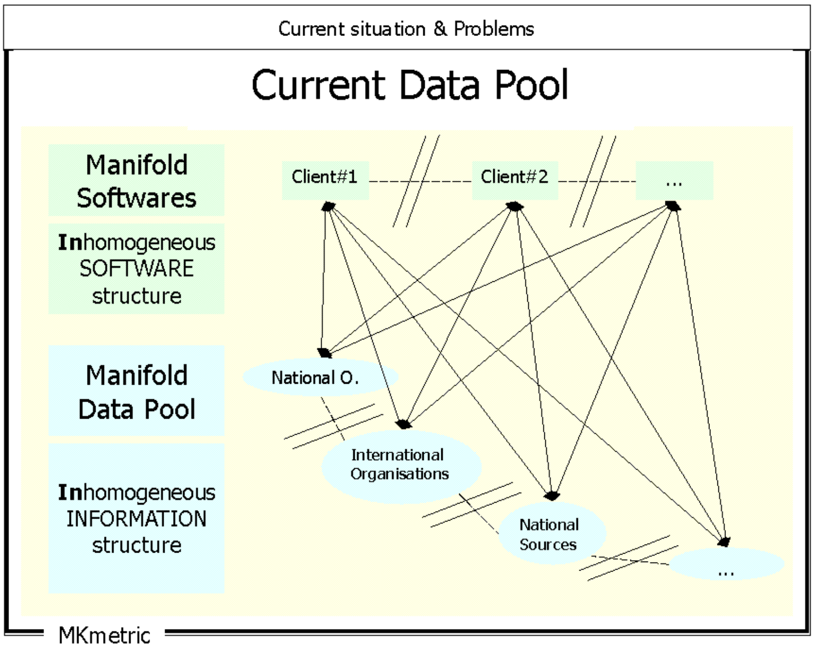


Fig. 3. Current Data Pool

The problem of trying to compare results from different software applications and databases lies in the fact that the underlying data-models (DM) and information structures for the model software applications and databases are incompatible or only made compatible with extreme difficulty and loss of information. This is because the informational units used by the data-models are to “large”, meaning one unit contains too much information. For example, the variable “time” is a typical variable used in (transport) modelling. But what does “time” mean exactly? The exact meaning is normally only known to the modeller himself and when this data (information) is transferred elsewhere, the problem of the exact definition of this variable arises, because “time” can mean “total travel time”, meaning for example for an air mode: network access / egress time + taxiing time + flight time. Mostly this information is implicit to the variable “time” used. What is needed is an explicit handling of this information.

Matching Data Models

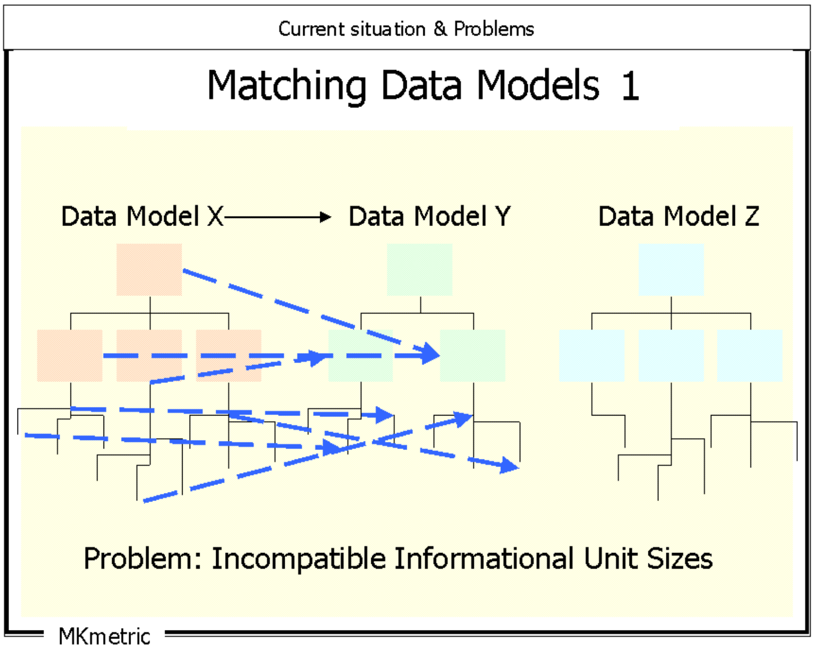


Fig. 4. Matching Data Models 1

Due to the “large” sizes of the information units, it is virtually impossible to manage the task of matching informational units contained in different databases without loss of information.

When more than one matching of informational units in different databases is required, there is a loss of information in each step of transformation, which very often leads to the fact, that the output of all the transformations won’t contain all expected and relevant information, see Fig. 5.

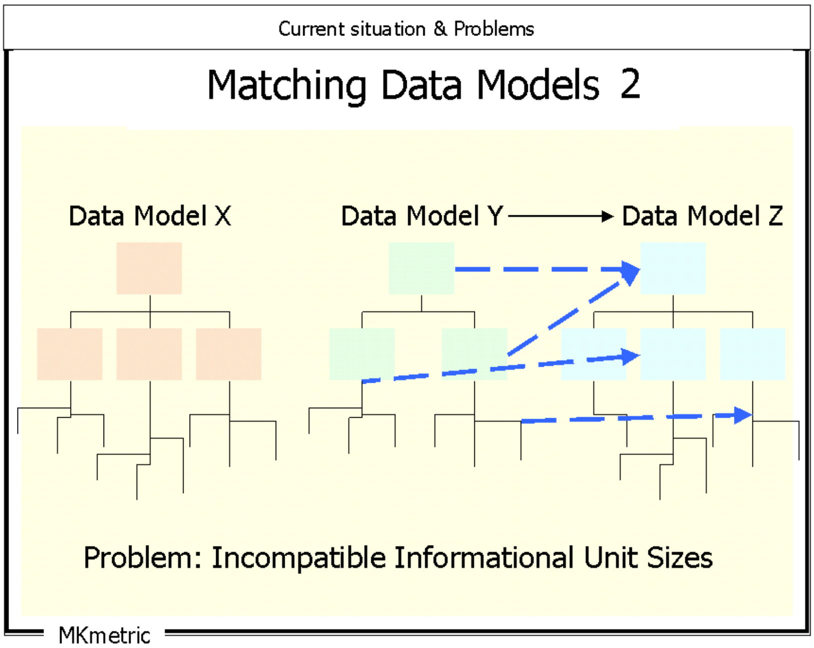


Fig. 5. Matching Data Models 2

2.5.1.2 Models & Data

A broad look at information & data for transportation models

One needs to examine the kinds of data and information used by different transport models, to approach the task of specifying an information / data model for GTF. Very generally speaking, transportation models use the following information items for their computations: 1. zonal data: any kind of zonal description, e.g. socio–economic data, ecological data, zonal boundaries, transport data, indicators, transport matrices etc. 2. network data: data describing the relations between the elements, e.g. link characteristics, a link has a starting node and an ending node (i.e. topological characteristics), link/network clusters etc. 3. GIS data: the necessary information for visualisation purposes, e.g. the underlying projection of the node and its co–ordinates. All these types of information must be must be part of the data exchanged with GTF. With these different kinds of information in mind, a more detailed view of the information / data categories for transportation information can be developed.

Typical Problems

Models in general are very demanding on the amount and quality of input and calibration data (e.g. even if they aren't discrete choice models). The main problems with current data and databases are 1. data required by the model, e.g. for estimation, isn't available. For example, a pan-European passenger transport model requires homogeneous input data (i.e. equal in structure and information contents) from all countries at the same aggregation level. This kind of database isn't available currently and when data (or interesting information for the model) is found not only the format does not correspond to the other data, but also the information contained in the new data lacks essential parts. I.e. the description of the data sounds good, but in detail there's missing a vital piece of information (or data). 2. the composition of the available data required by the model doesn't match and re-composition isn't possible. For example, a database that was acquired for a model, has the data in an aggregation that can't be matched to the one needed by the model. If the model requires a NUTS zonal division [EUROSTAT94] [EUROSTAT95] of the data but the acquired data has a different regionalisation (e.g. ITP) the units of data (in the first case a NUTS zone and in the second case an ITP zone) do not match exactly, thus re-composition of the ITP data into data fitting the NUTS regionalisation is necessary. This can only be done, if further information concerning the amount of the ITP data for a zone that fits into the equivalent NUTS zone (i.e. the percentage of the ITP data for the zone that fits into the NUTS zone) is available. Also, it is usually the case, that you need to take different percentages from a set of ITP zones to create an equivalent NUTS zone (and vice versa). Making the re-composition a tedious and error prone task. 3. the aggregation of the available data doesn't match. This often means, that the acquired data is aggregated to a higher level than required and can't be disaggregated to the level needed, because e.g. the disaggregation procedure isn't known.

Conclusion

Because of the problems mentioned, there is the potential of significantly increasing the value of models' databases by not only homogenising them but by defining a generalised structure with which information contained in (current and future) databases and the required information (not just the data) for models can be described precisely [SPOTLIGHTSTN]. The first (and main) advantage would be to have databases which can be exchanged, enriched, corrected and used by models in a homogenised manner, drastically reducing the amount of redundant work, e.g. by having to reformat and re-arrange pieces of data in order to fit the data requirements of a specific model. Secondly, it can be ensured that the required information is actually contained in the data.

The previous sections, gave a brief overview of the problems that arise when having to do major re-formatting tasks. Therefore, a “Generalised Transportation-data Format” Data Model will be developed in the following sections. Also a specific exchange format will be defined to be able to process (exchange) concrete GTF Data Models by electronic means.

2.5.2 GENERALISED TRANSPORTATION-DATA FORMAT (GTF)

Basically, the GTF data model is a framework which can be used to define the information that is contained in data. The GTF data model framework allows a user of the GTF specification to wrap data into information entities. These entities contain the basic data and the necessary supplementary information (meta–data) to give a meaning to the basic data. Like this, one can make sure that the input data to a transport model fits the model’s information requirement. This is vital for a model to compute valid results. If the input to a model doesn’t fit the assumptions that were made concerning the information carried in the data (i.e. the meta–data implicitly associated to the input data), then the model is very likely not going to produce valid results. Here, the GTF data model comes into play. It enforces a user of this specification to make the implicit information explicit by wrapping the data – data with implicit information – into entity structures – data with explicit information. These entities are then combined to represent the complete implicit information that a piece of data carries. In this way all the data’s information is made explicit and thus can be used to check whether data fits the model’s philosophy or not.

How can one know which entity to use to wrap data into? The first step is to make it clear to oneself which information a piece of data is carrying. For example, a piece of data could be “60”, but the information it carries for a user of the GTF data model (typically a model provider that needs to package the model’s results into a GTF-XML file according to this specification) is “speed”, “maximum”, “kilometres per hour”, “administrative regulation for zones in Germany” and “on the Link 87321 between Node 983 and Node 1001”. Now, one could define a new and separate attribute “maximum speed in km/h for Germany” for the Link between Node 983 and Node 1001. But to do this for every possible combination of these pieces of information is not possible due to the combinatorial explosion. It would also be very ineffective to define always new attributes for a piece of new information. The question is also, what happens if a new piece of information is associated to the data, e.g. “for cars not for trucks”. Shall a new attribute be defined? This doesn’t seem very elegant.

Therefore, the GTF data model provides a standard set of information pieces that can be used to wrap data into and to combine the pieces of information into a larger chunk of information, just like building something using LEGO[1]. With LEGO one can build many different things without having to buy always different pieces of LEGO. One can use the standard set of LEGO pieces and still build a very large number of different things. One only needs a really new piece of LEGO, if one wants to build something that doesn't fit into the concepts of the available LEGO pieces. For example, with only square LEGOs it is impossible to build something round (– at least it's very difficult). In this case, the concept of "round" is totally different to the concept of "square". The "square" concept is covered by the available pieces of LEGO, the "round" concept is not. Thus, one uses a new "round" LEGO to bring the concept to life. In the case of GTF, all this applies equally, the different concepts of LEGO pieces, e.g. "round", "square", are defined as entities. Each specific real piece of LEGO is an entity instance. A construction made of LEGO is a GTF-XML file with the entity instances and the definitions of the relationships between them. The main advantage of this kind of thinking is the minimal amount of different abstract concepts used to cover a very wide range of concrete things.

Coming back to the question "how can one know which entity to wrap data into". The answer is fairly simple. Think about what information is implicit in a piece of data and choose the correct combination of entities from the data model and follow the pre–defined relationships of the data model.

The GTF data model is a complete data model in the sense that all parent entities required to define a child entity are also defined as separate entities, although the parent entities are abstract and mostly mentioned to have a compete framework. The entities that are actually used in a GTF-XML transmission have a definition of a XML segment following the definition of the attributes associated with the entity. For example, the entity Terminator which only captures the concept of "something that is the beginning or the end of something else" is abstract, the concrete concept is an intersection Node which is "a point in an infrastructure network". Thus, in a GTF-XML only Nodes should be transmitted, because the Terminator information is automatically known. This information is implicit in the GTF-XML file, but explicit when looking at the Node definition in this specification. This means, that when a Node is transmitted, the receiver (who also has to know this specification) automatically knows that a Terminator is the parent of the Node. One can transmit a Terminator, if needed, but a

[1] LEGO is a trademark and copyrighted by LEGO Systems AIS, Denmark

Terminator isn't concrete and can, therefore, be either a Node or a Zone. It isn't possible to determine which, according only to the Terminator information. The information whether a Terminator's role is an intersection Node or a centroid Node is explicitly contained in the corresponding entities in the data model. Thus, one Terminator can be used as parent of an intersection Node and a centroid Node, if the information to be conveyed is: "This starting / ending point has the role of an intersection Node and it is also the input / output point of a zone, i.e. the zone's centroid, in this network".

One might be flabbergasted by the sheer number of defined entities (approx. 200) when looking at the list of entities in the data model. But the basic entities that were used to create the data model are a total of only 9, namely SpawnFactor, Terminator, Link, Specification, Vessel, Service, Alternative, Unit and Meta which are called "topmost entities" or "top–levels".

The top–levels and their children (called "concretes") can be combined using the defined relationships. These relationships are defined by a user of this data model by filling out an attribute in an entity that was migrated to the entity through the relationship. This means that the entity attributes defined in the data model are either generic to the entity or have been added to the entity because of a relationship.

Once it is clear which entities to use for the transmission of a piece of data, one needs to generate the corresponding GTF-XML segments and construct a complete GTF-XML interchange transmission file.

2.5.2.1 Definitions

More specifically, what is meant by "Transportation Entity" in this paper is:

1. Transportation = "The act of moving passengers or freight in space."
2. Transportation Entity = "All that spawns, enables or hinders movement of passengers or freight."
3. Transportation Relationship = "The connection between two transportation entities."
4. Transportation Attribute = "A quality or feature of a transportation entity that is a central part of its nature distinguishing its instances."

For example, concerning the entities Zone & SpawnFactor, this means, what is meant by SpawnFactor are those information that are not essentially part of a Zone's nature, e.g. age distribution etc.

The things that are part of a Zone's nature are attributes of the entity, e.g. barriers (mountains, lakes etc.) separating one Zone from another. This kind of information is kept within the Zone entity. Also, the definition of SpawnFactor in GTF not only contains the raw data, but also the meta-data. I.e. each SpawnFactor instance contains the raw data, e.g. 5, and the necessary meta-data, e.g. “statistical source = EUROSTAT, type = statistics”.

Obviously the definitions above cannot be used for direct implementation. The goal of these definitions are to be able to define a data model of transportation and not to implement this data model. The implementation is left to eventual providers who have to adopt GTF as one of the exchange formats of their software/model.

These definitions provide a good and general basis to derive a data model for GTF. Now reading these definitions one might argue “since a pile of leaves might hinder movement, by definition it should be an entity of transportation. Does that make sense?” Basically, yes, it does make sense. Although practically, no. Since GTF was conceived as a data model for (mainly) strategic transportation models which are long-term (e.g. 10-50 years forecasting flows, impacts etc.). If say one would devise a model that takes “piles of leaves” into account, then it is an entity instance (i.e. an object from a class [BROWN97] [BUDD97] [RUMBAUGH91]). And in fact, the GTF data model does contain objects called “SpawnFactor”s that are used to store all the spawn-factors which generate, attract (basically make or influence) movement and the data model also contains the entity “Barrier” (associated with Zone) to describe all objects that are in the way of free movement (mostly, e.g. mountains, rivers etc.). So the answer is, “if the pile of leaves hinders movement, say on Link L, for the totality of the period being analysed with the model, then yes, these leaves have to be considered an entity of transportation, just like you would consider a river flowing between two Zones, also being a transportation entity which hinders movement.”

2.5.2.2 Structure Overview

Due to all the problems described above, the structure of a “Generalised Transportation-data Format” data model should cover the following 1. instead of having disparate and manifold software applications and databases, GTF contains all necessary elements and provides one single and homogenous data specification and format 2. instead of having incompatible proprietary formats and informational contents, GTF should be used throughout any computer system, by providing translators to / from the proprietary formats to GTF.

GTF consists of 1. a generalised data-model (GTF-DM), 2. a standard exchange format (GTF-XML) and 3. generic commands to run models and retrieve results (TIP).

GTF Data Pool

With GTF, the structure of the numerous software applications and databases are accessible in a homogeneous and compatible manner. A set of GTF Translators could provide a single access point to all models and data. The problems discussed previously of inhomogeneous software and data / informational structures and definitions is overcome by using the GTF Data Model specification [MKMETRIC99] [EUROSTAT96] to structure and flesh-out databases and for the exchange of information (by using GTF-XML), see

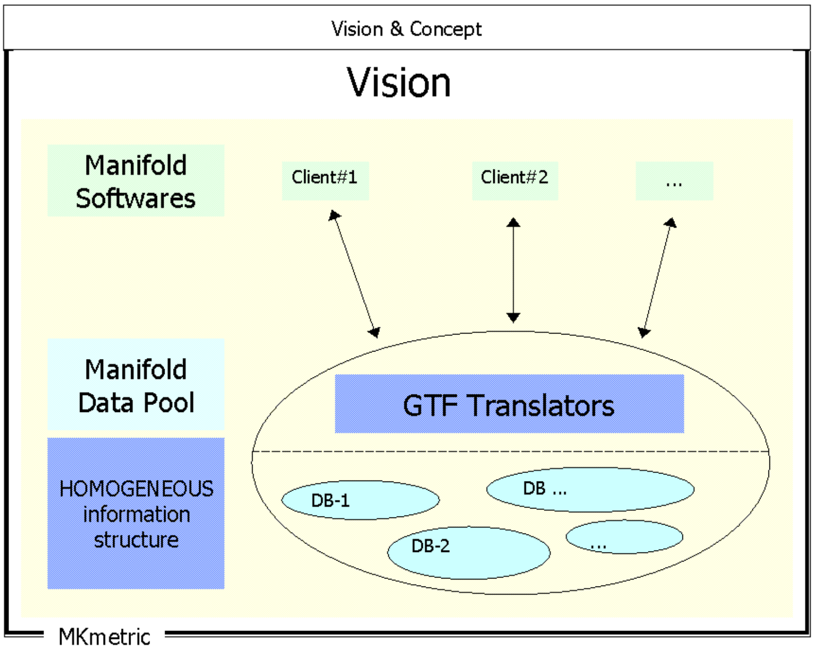


Fig. 6. GTF Data Pool

The numerous databases can either be restructured according to the GTF Data Model or a specific GTF Translator for each database can be developed thus providing a homogeneous and single access possibility (see Fig. 6). Software like GIS can be enhanced by a GTF Translator to be able to access the information of GTF databases.

Matching Data Models using GTF Step 1

The main concept for the development of a GTF Data Model is that the informational units of GTF are "atomic". Therefore the informational units (- the data) of any other DM (DM-X) can be decomposed according to the GTF-DM (see Fig. 7).

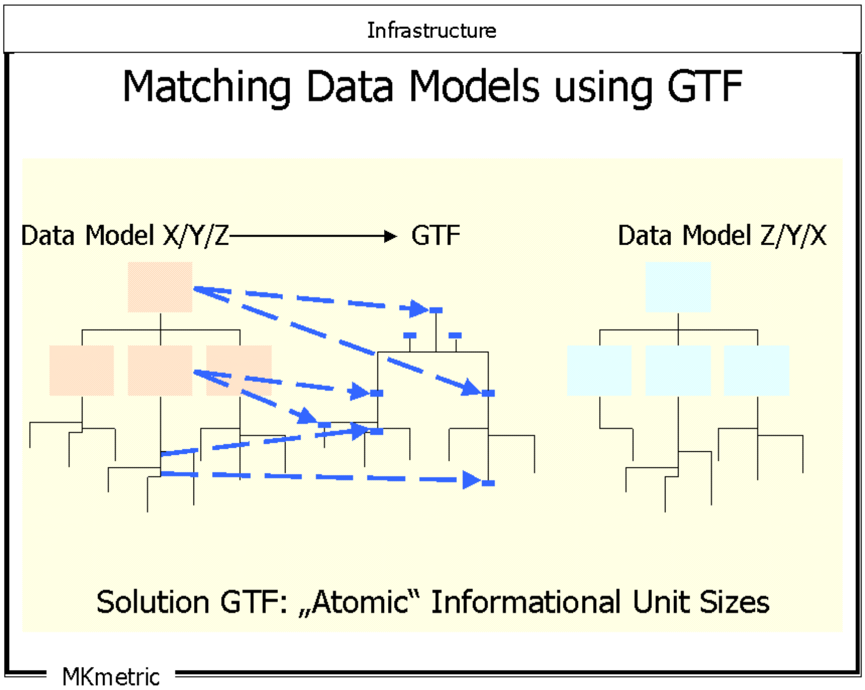


Fig. 7. Matching Data Models using GTF step 1

For example, an information unit of some data-model DM-X of "flows", say "time", is "50". This is a non atomic information because one can't know whether there is much implicit information or not, without having the exact definition of the data-model (and thus the exact definition of the attribute "time"). "Time" might be an aggregated value meaning the "Total travel time". This, too, is aggregated and could mean f.i. the sum of access / egress, taxiing time, flight time, etc. Already this example shows that implicit information can be numerous. To avoid this confusion, the GTF data-model offers containers of "atomic" information which must be used to describe aggregated information as in the example above.

In the GTF-DM, all pieces of information that qualify a piece of data are kept in separate entity instances which are linked through relationships to the entity instance containing the piece of raw data. In the case of the example, this means, that "access/egress", "flight time", "total time" etc. are qualifiers ("flags") attached to the raw data "50". In this way, the implicit information is made explicit, because each implicit piece of information is reflected in an entity in the GTF-DM.

Matching Data Models using GTF Step 2

Following the example in the previous section, the decomposed information from DM-X can be re-composed differently than in the original DM-X using the GTF informational units, creating the data according to a different DM (DM-Z). I.e. assembling of aggregated information units from the GTF-DM units can be done in analogy to the decomposition of aggregated information units into GTF-DM units (see Fig. 8).
The main focus for the development of the GTF specification and subsequently the GTF Translators is

> *to define an abstract view of transportation models' information for the purpose of implementing translator software to exchange data electronically between modelling software and other software (e.g. model packages, database systems).*

The primary goal therefore is the definition of a data model for transportation information (a GTF specification of information structure) and the definition of a format and syntax for electronic transfer of information (a GTF Translator syntax and the format of information data).

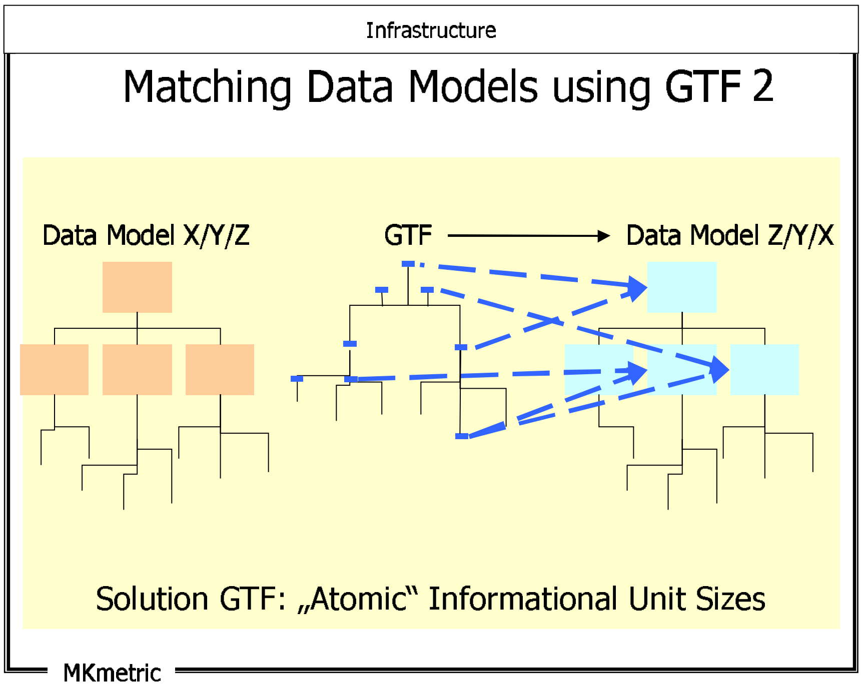


Fig. 8. Matching Data Models using GTF step 2

2.5.2.3 The GTF Data Model

This section introduces in more detail the fundamental information classes that are the foundation of the GTF Data Model. The transport data that is covered is primarily that which is used in strategic transport models. Thus it covers interurban, regional or international travel on all transport modes for both passengers and freight. It does not cover detailed local traffic issues, such as the representation of road junction geometry [MOELLERING97]. But this can be described using GTF, too.

GTF Data Model Overview

Some basic concepts from economic theory (i.e. supply side determinants, demand side determinants and the market where supply meets demand) were used to develop the concepts for the data model [BUTTON93] [ORTUZAR90]. Fig. 9 depicts the main conceptual entities used in modelling information (without going into details).
A number of SpawnFactors (not depicted) determine the generated or attracted movement, which together induce the demand for movement and transport. SpawnFactors are for example the GDP, age distribution, level of income etc. for one Zone, a group of Zones or an aggregation of Zones.
A centroid Node is a virtual point for input & output (source & sink) of movement in networks. It is associated with Zone which contains the SpawnFactors of an area. For transportation models a Zone is a description of socio-economic and other information of a geographical area. The geographical connection between a Zone and the area it describes is used to relate specific SpawnFactors and their values to specific input and output points in infrastructure networks. In this context the virtual input & output points are called centroid Nodes. These kinds of Nodes are not to be confused with infrastructure (network) Nodes, that are descriptions of junction points which in turn are an abstract description of some part of a physical transportation network.

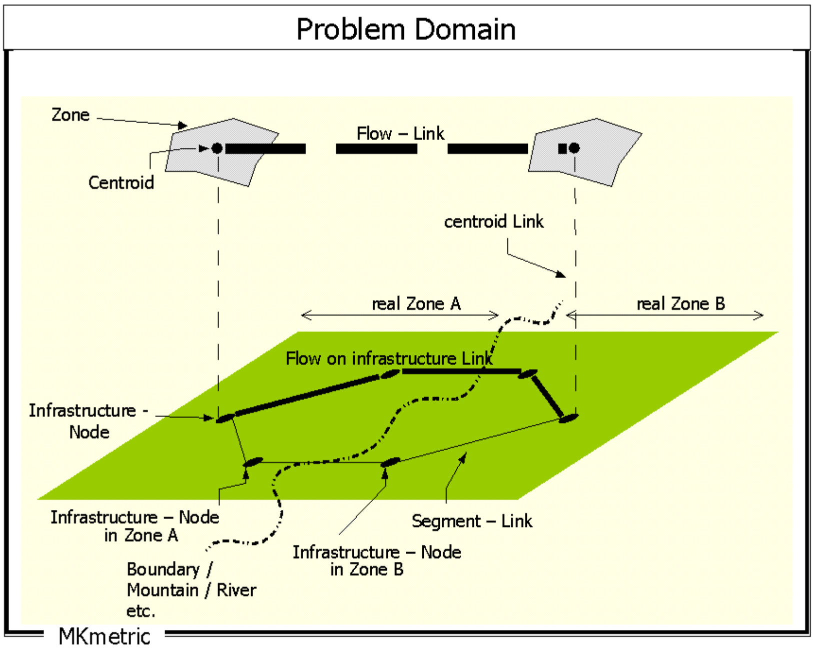


Fig. 9. GTF Data Model Overview

The concept behind SpawnFactor and Zone is the following: an area itself (e.g. 10 km^2) cannot be the reason why traffic or movement is produced or attracted. The reason why there is movement to and from an area, is that the area has a number of features that the travellers to an area are interested in. In this paper these features, e.g. people, industry etc., are called SpawnFactors in analogy to the concept of production factors in economic theory. A Zone then, is the combination of an area (either in physical space or in the modelling space) and the SpawnFactors, that are located in the Zone's area. The specifics of the Zone's area are described by the combination of SpawnFactors and a Zone represented by the relationship "activity", because a SpawnFactor in a Zone generates some sort of activity, either attracting movement or producing it.

A centroid is connected to an infrastructure network through a centroid Link. The centroid Link is the virtual description of the impedance(s) that is needed in average to enter / leave a Zone and thus creating inter-zonal transportation / movement called LinkFlow.

A LinkFlow is the result of SpawnFactors generating & attracting movement across the limits of Zones. It can therefore be described as a connection (relationship) between two Zones. This is "flow" in the sense of demand for transportation. A LinkFlow is the container of information that exists when two specific Zones are connected, because there is demand for movement between them. Thus, a LinkFlow is a connection between centroid Nodes with information about the amount and types of flows (vehicles etc.) that go into and come out of a Zone.

A “node” performs two functions in transportation modelling. The first function is to relate (connect) a Zone to some point in the network as access and egress points for mobility. This function determines the Node as being a “centroid Node”. The other function is that of being a junction point, which determines the Node as an “intersection Node”. These junction points describe topological aspects of infrastructure networks, i.e. which Nodes are connected to which other Nodes. The Nodes and the connections (branch, arc, edge etc. called Link) between them are the topological description of the networks.

A Link is a topological relation between two Nodes. The Nodes in turn usually are associated to specific geographical co-ordinates in real world space. But this is mostly needed for visualisation and presentation purposes.

An “activity” is a term for “determines demand of”. This term was chosen, as it describes better the concept of the entity SpawnFactor. “activity” can also be seen as abstraction of the attractiveness of Zones or the potential for a visitor of the Zone to see sights etc. An “activity” describes everything that induces movement/transportation to/from a Zone.

Following this kind of logic of decomposition and analysis, 9 toplevel[2] entities are defined in the GTF: Terminator, Link, Vessel, Service, Alternative, SpawnFactor, Specification, Unit and Meta. The next table gives definitions for each toplevel.

Entity Name	**Description**
Terminator	This entity is the generalisation of the concept “start or ending point of Links” and thus a generalisation of the centroid and intersection Node concepts usually used in modelling and graph theory; and its function is to act as the starting / ending point of Links. The Terminator entity is the generic entity for both types of Node. The generic entity Link therefore is determined by exactly two Terminator entities which can be either a centroid or an intersection Node. In this way a more homogenous view on the problem domain and especially the similarities between centroid and intersection Node with respect to their function defining Links is achieved. FUNCTION: the starting / ending points of Links

[2] A toplevel entity is a parent from which a hierarchy is derived. The toplevel entities and their relationships are the foundation of the complete GTF Data Model.

Entity Name	Description
Link	The Link entity is not only an abstraction for all types of infrastructure network links, but it incorporates the connections between two Zones (through their centroid Nodes) when modelling flows. Thus, a (zonal) Link is specified by exactly two centroid Nodes (of two Zones) which are the starting and ending points of a flow. Centroid and intersection Nodes in different combinations act as Terminators to define Links. The three possible types of Link are (depending on the combination of centroid and intersection Node): 1. the LinkInfrastructure is a Link between two intersection Nodes that is used to describe the supply-side of transport, i.e. infrastructure elements that supply the possibility of movement / transport, e.g. road links etc. 2. the LinkConnector between an intersection Node and a centroid Node is a Link that describes the avg. travel-times, costs, speeds describing the avg. disutility to reach (any) point in the Zone. 3. the LinkFlow between two centroid Nodes is a Link that holds the flow information that results when two Zones are connected to describe the movement between two areas in space. FUNCTION: the possibility of movement between two Terminators
SpawnFactor	SpawnFactors determine the generated or attracted movement of a Zone, which together induce the demand for movement and transport. SpawnFactors are for example the GDP, age distribution, level of income etc. for one Zone, a group of Zones or an aggregation of Zones. A SpawnFactor is a piece of data (aggregated or disaggregated), e.g. socio-economic or other statistical data, that is used to compute / describe the potential for transportation demand that an actor / group of actors generate / attract. A SpawnFactor is unique to a Zone, because a SpawnFactor has an explicit value for some demand SpawnFactor of the Zone, e.g. GDP=5000, which is Zone specific. SpawnFactor objects can be seen as a container of the attributes of the associated Zone object. Because of its importance it was defined explicitly as an own class. FUNCTION: it is used to describe actors (or group of actors) attributes that are used for Zones in the transportation model. Actors are the reason why movement and flows are generated or attracted.
Specification	Specifications are characteristics which can be attributed to Links. The Specifications of a Link can be defined by Zones, Vessels and Units. For example, information concerning speed-limits or tolls are defined by a Zone’s location and its political / administrative regulations. The Specifications can be dependent on a

Entity Name	Description
	Zone's country attribute value, because of the relationship "defines regulations for". These specifications are regulatory / administrative or defined by engineering science. The interesting point to note here is the fact that the "specification" (e.g. maximum-speed limits from an engineering point-of-view) of a Vessel like a car can depend on the Zone's location (e.g. Germany, UK) and the type of Link (e.g. motorway, 2-lanes, 4-lanes), because the law and regulations for the different types of Vessel that are allowed to use different types of Link differ by country. FUNCTION: this entity associates all the technical, statistical and movement specifications that come from Zone, Vessel and Unit and defines (physical) characteristics of a Link
Alternative	Models use choice alternatives (e.g. usage of road or rail or air mode for transportation etc.) to describe the situation individuals (or the behavioural units being modelled) face in certain situations. The model then "decides" which option the individual chooses by taking into account different aspects (socio-economic, economic, psychological etc.). From a modelling point-of-view the Networks (i.e. the groupings of Nodes, Links etc. which form a logical whole) need to be distinguished according to different "main modes" (or Alternatives), because models use these "main modes" to differentiate elements of sets of choice alternatives. As the term "main mode" intuitively implies one single mode (i.e. Vessel), which isn't the proper concept, because a model's "main mode" can imply any number of modes, the more precise term Alternative, short for "choice alternative" is used. The instances of the Alternative entity define choice alternatives (or choice of combinations of available means of transport) for a model, through the combination of Units (person/good/business/private/holiday ...), Services and Vessels. These choice alternatives are "main modes" when associated to Links. Certain Links can only be used by some of the Alternatives, because e.g. the Link can't cope with a Vessel of some given tonnage used in the Alternative definition or some other restriction due to the definition in the model, i.e. an Alternative's main mode might not be allowed to use a Link, but otherwise (physically) it is allowed. Note, the term "main mode" is an alias for "choice alternative" and doesn't imply only one single mode. Because an alternative can be defined using any number of "modes" (Vessels). The entity Alternative comprises all the definitions for a Link that are derived from the modelling-side of the information, e.g. the "main mode" might be "road". This actually comprises road

Entity Name	Description
	links as well as car ferries etc. The Alternative entity associates this information to a Link. The difference to the entity Specification (and Vessel) is, the Alternative describes a logical (modelling) use of a Link while Specification describes the “real world” usage of a Link, e.g. a Link might be a ferry link, but if it is a car-ferry the modelling usage would say that the link is a road link. FUNCTION: a container for information pertaining to the definitions of choice alternatives for a model
Service	A service provides a traveller with the means to travel with relevant choices already made in advance by the service operator. The Service entity is a container for information pertaining to services, e.g. public transport. This entity is a definition of a type of service, the used carrier Vessel(s), the level of security attributed to this type of service and the time-table for the service. The instances of Service are used by Link and Alternative. The relationship “real definition” to Vessels, associates the Vessels that a Service uses to support their service. The entity Service is the container for specifications concerning services, carriers etc. that use a Link. With this entity a Link also has “usage” information apart from the “physical” usage. FUNCTION: bundling of assistance to a user (=traveller) for travelling purposes
Vessel	Vessel is the abstraction of everything that increases the flow-count on any Link. In transportation models typical Vessels are cars, trains, aeroplanes, trucks etc. A vessel is a logical view of objects / entities that can use links to travel / transport some person / good from one point (Terminator) to another. A vessel description contains all that characterises a vessel object. A Vessel is anything that moves and uses infrastructure entities, e.g. cars, planes, persons that use roads, rails, airways etc. There is also the virtual Vessel like a “human” that uses the mode “walking”. For example for the access / egress points (where the car is parked) of a railway station or airport to the points where one actually enters the railway station or airport, one has to walk. Thus one is using the mode “walking”. FUNCTION: a container of information of that that travels / moves on Links
Unit	Units define the type of unit being moved or transported, the purpose of the movement or the date / time schedule of a movement. This entity contains all such information and associates this information with Alternatives and Specifications etc.
Meta	Metas are objects to define meta-information which do not pertain to modelling or

Entity Name	Description
	network or other information, but are rather complementary information describing units of measurements etc. The Metas can also be used to associate dimension information with all other entity instances, because each Meta instance has two attributes “for entity code” and “for instance”, which uniquely associate the Unit instance with the instance of some other entity (with that entity code).

Using these high-level definitions approx. 200 entities are defined in the data model.

It is important to note that an entity instance’s function in the data model is either to actually hold raw data or to serve as a qualifier to another entity instance holding the raw data.

Fig. 10 depicts the toplevel objects and their relationships.

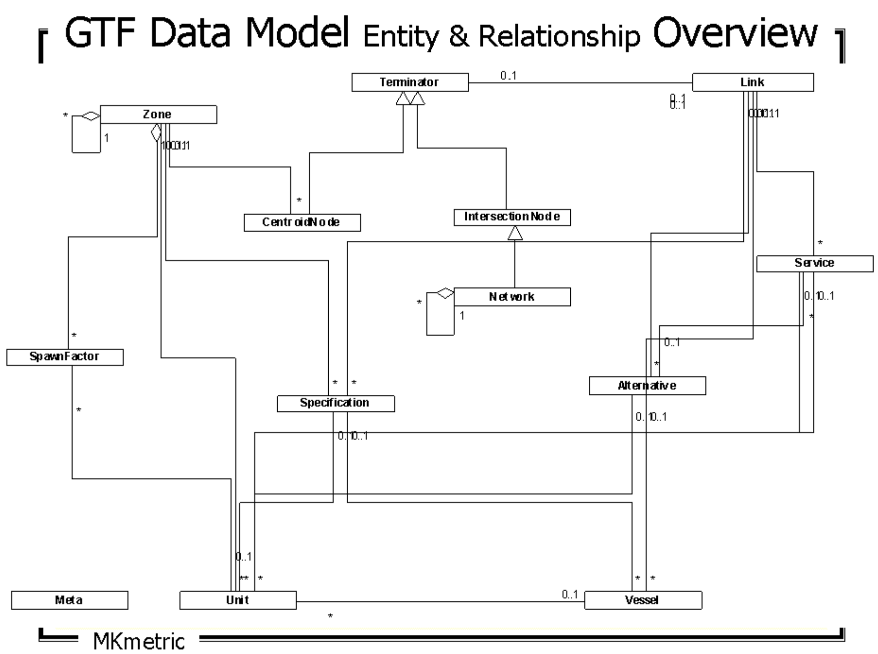


Fig. 10. Overview GTF-DM

These basic types (entity or class) of information are further sub-divided in the GTF-DM until the level of detail required (one bit) is reached. Fig. 11 [UML] gives a bit more detail about the data model structure at a high-level.

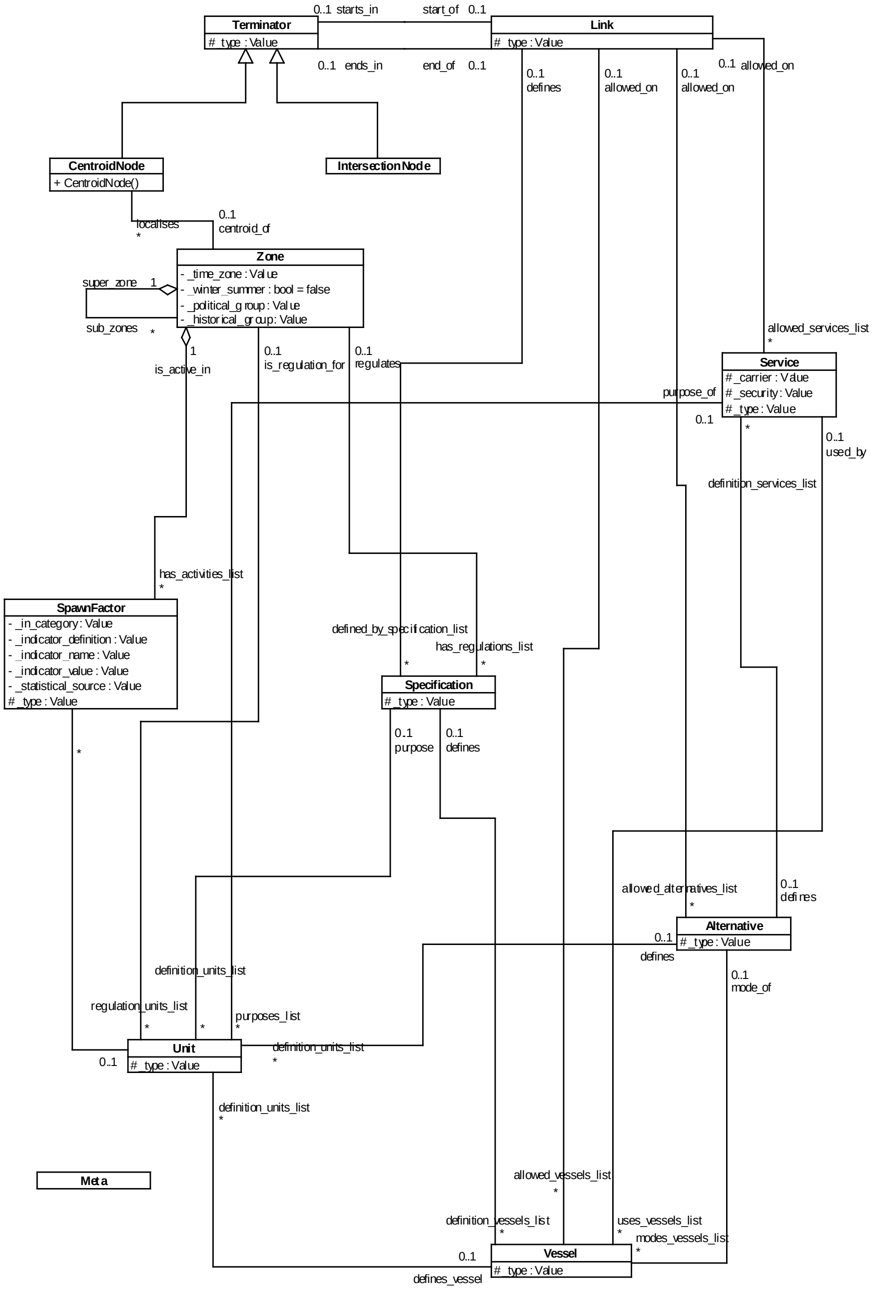


Fig. 11. UML Diagram of toplevel Entities & Relationships

All entities of the GTF-DM have a “GIS” part and a “TYPE” part. The “GIS”-part is used to capture graphical information, e.g. co-ordinates, vertices attached to Links etc. The “TYPE” part contains the information relevant to models, e.g. modes that are part of an alternative. To explain the principles of the data model structures as an example, the informational unit “Terminator” structure is depicted in Fig. 12.

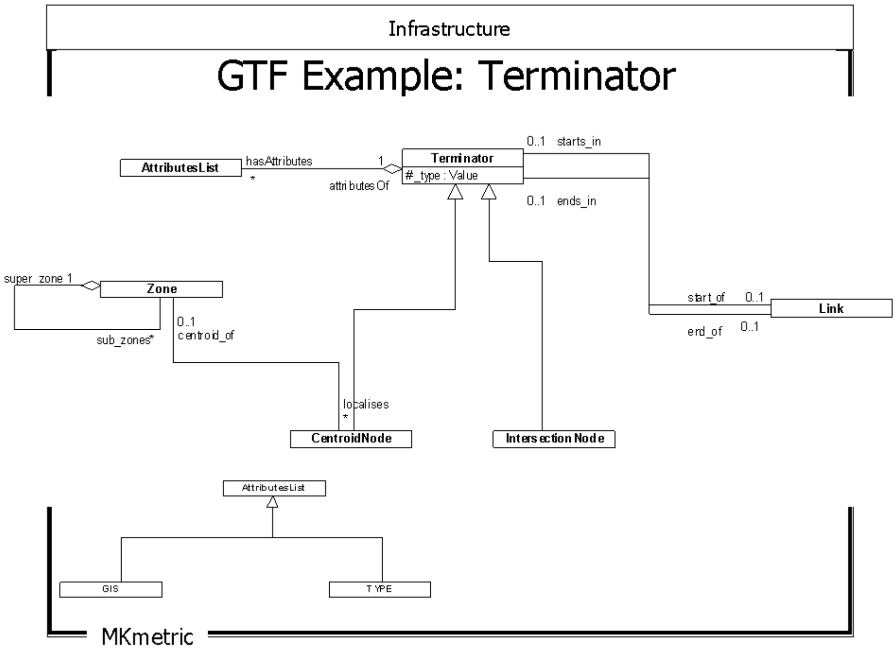


Fig. 12. GTF Example: Terminator

All Terminators (all entities) have a Type entity of information and (optionally) a GIS entity of information associated to them.

The GIS entity information of a Terminator is for example Symbol entity = definition of the symbol to use when displaying the Terminator graphically in say a GIS. This entity has a relationship to a Shape entity which captures the co-ordinates defining the Shape.

The Type entity information of a Terminator are for example Infrastructure entity, defining the Terminator as a network intersection Terminator without further structure. (In contrast, the Type being (sub-)Network entity would define the Terminator as a sub-network, which itself is a grouping of (possibly) all other entities in the GTF-DM. This allows for different levels of detail in a network.)

All Terminators have a relationship to a Zone (called "is in", not depicted), defining the Zone in which a Terminator is contained. All Zones have relationships to centroid Nodes (called "localises") defining the centroid Nodes which connect the Zone to the network. Intersection Node and centroid Node are children entities of the Terminator entity. The function of Nodes is to represent (in combination with Links) the graphical and modelling topology of networks, the Zones to group factors of activity.

For example, the Terminator–Node–GIS entity is the container of graphical Node information like projection, zoom–level of the co–ordinates associated with Node. The

entity Terminator–Node–GIS–Symbol-Shape is a container of lists of co–ordinates describing the shape of a symbol used to display Nodes.
The function of a Node as an element in the topology description of a supply side model makes a Node of type “infrastructure”. If the Node is used as an aggregation container of other entities, then the Node is of type (sub–) Network entity. And one uses this kind of Node to “zoom in” and to “zoom out” of a Node in order to see its internal structure. The concept of “zooming in” is to show further **topological** details (not only graphical details) associated with the Node. The further detail a (sub–)Network can associate with a Node is, that the Node is made up of other entities, e.g. a group of Nodes and Links that describes a railway station, an airport or generally terminals and their access and egress points as well as their “turns” and “changes” between Links (or Links of different modes).

2.5.2.4 Fundamental Design of GTF Translators

Requirements

From the description of the requirements of the system follows that modelling-data needs to be transferred across different platforms, mainly Windows and UNIX platforms. This is because many modelling software applications are implemented on UNIX platforms and the default platform for users is (usually) a PC.
The structural system requirements [MKMETRIC/MESUDEMO99] are depicted in Fig. 13.

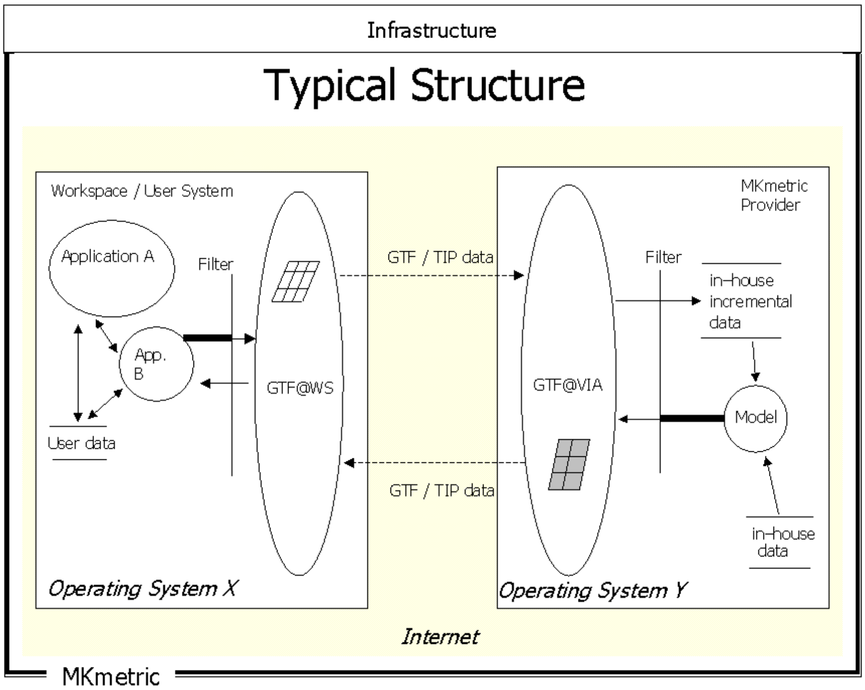


Fig. 13. Typical Exchange Situation

A user modifies his local “User data” through his system. He then formulates a request for a model. The user specifies the model to use. The user also specifies the request and the data to be used. A filter is used to make sure that only relevant data (or data not unknown to the model provider) gets translated by the GTF@VIA Translator (VIA = MKmetric’s model package). The resulting GTF file is transferred to the user’s account at the model provider’s server. There the data from the GTF file is extracted and incorporated as incremental data into the data already available at the provider’s (in–house data) site. The complete data is then fed into the chosen model according to the TIP information in the GTF file and the requested computations are done. The requested results are extracted by the filter and translated into GTF by the GTF@VIA Translator. The user’s system gets notified that the requested results are ready for download at the provider’s site. The user downloads the data. The user can then view the results with his favourite applications.

The consequences for the actual structure of a GTF file are:

Cross-platform / Human-readability

A non–binary code must be used. The choice has fallen to the ASCII code, because this format has the least problems when being exchanged between heterogeneous platforms. ASCII also has the additional effect that a GTF file in ASCII can also be read and understood by a human, in case of problems.

Segmented & Self-describing

As the data and control information to a model needs to be put together by the user's system the exchange format must be very flexible and powerful. The best way to achieve these two goals is to design the format and protocol for interchanging in a structured and segmented file. Like this, the user's system has a "language" to describe the structure and contents of the GTF file.

2.5.2.5 GTF data model & GTF–XML message specification

Ad-Hoc-Format

A simple ad-hoc format is defined based on XML [MARCHAL98] to be able to describe some examples of how to use the GTF data model.
This is an ad-hoc format, because the development of the GTF specification didn't have the goal of defining a universally accepted computer exchange format for the data model. This ad-hoc format was defined in order to make concrete examples of how to use the data model. It is based on XML, but isn't a validated and formally correct XML format.
The next two lists define 1. the elements of the format 2. the organisational structure. It is assumed that the XML format and syntax are known. (An introduction to XML can be found at [MARCHAL98].)
List of toplevel entities for the GTF-XML:

Block-level Entity (Object)	XML Tag	Attributes	Comment
GTFDB	<GTFDB> </GTFDB>	id, name	Top most element. Hierarchical starting instance of the information network using the other entity instances.
Terminator	<T ></T>	id, name	Default attributes
		start_of, end_of	Ids of the Link entity instance of which this Terminator is the start / end of
		type	value from code list
Link	<L></L>	id, name	Default attributes
		starts_in, ends_in	Ids of the Terminator instances which are the start / end of this

Block-level Entity (Object)	XML Tag	Attributes	Comment
			Link instance
		defined_by_specification_list	list of Specification entity instance Ids
		allowed_vessels_list	list of Vessel entity instance Ids
		allowed_alternatives_list	list of Alternative entity instance Ids
		type	value from code list
Zone	<Z>	id, name	Default attributes
		time_zone	value from code list
		winter_summer	value from code list
		historical_group	value from code list
		political_group	value from code list
		has_activities_list	list of SpawnFactor entity instance Ids
		localises	list of centroid Node entity instance Ids
		super_zone	Id of the Zone entity instance of which this entity is a part of
		sub_zones	list of Ids of the Zone entity instances which compose this Zone entity instance
		separated_by_barrier_on_the_left	Id of the Barrier entity instance
		separated_by_barrier_on_the_right	Id of the Barrier entity instance
Intersection Node	<N></N>	id, name	Default attributes
		type	value from code list
SpawnFactor	<F></F>	id, name	Default attributes
		in_category	value from code list
		indicator_definition	value from code list
		indicator_name	value from code list or TEXT
		indicator_value	TEXT indicating the value
		statistical_source	value from code list or TEXT
		is_active_in	Id of Zone entity instance
		type	value from code list

Block-level Entity (Object)	XML Tag	Attributes	Comment
Specification	<SP></SP>	id, name	Default attributes
		regulates	Id of Zone entity instance
		definition_units_list	list of Unit entity instance Ids
		definition_metas_list	list of Meta entity instance Ids
		definition_vessels_list	Id of Vessel entity instance Ids
		type	value from code list
Alternative	<A>	id, name	Default attributes
		definition_services_list	list of Service entity instance Ids
		mode_vessels_list	list of Vessel entity instance Ids
		definition_units_list	list of Unit entity instance Ids
		definition_metas_list	list of Meta entity instance Ids
		type	value from code list
Service	<SE></SE>	id, name	Default attributes
		allowed_on	Id of Link entity instance
		uses_vessels_list	list of Vessel entity instance Ids
		defines	Id of Alternative entity instance
		purpose_list	list of Unit entity instance Ids
		schedule	list of UnitDateSchedule entity instance Ids
		type	value from code list
Vessel	<V></V>	id, name	Default attributes
		allowed_on	Id of Link entity instance
		mode_of	Id of Alternative entity instance
		definition_units_list	list of Unit entity instance Ids
		definition_metas_list	list of Meta entity instance Ids
		type	value from code list
Unit	<U></U>	id, name	Default attributes
		information	TEXT representing the actual data
		for_entity_code	TEXT, name of a class
		for_instance	Id of any entity instance of class “for_entity_code”
		type	value from code list
Meta	<M></M>	id, name	Default attributes
		information	TEXT representing the actual data
		for_entity_code	TEXT, name of a class
		for_instance	Id of any entity instance of class

Block-level Entity (Object)	XML Tag	Attributes	Comment
			“for_entity_code”
		type	value from code list
Network	<NE>	id, name	Default attributes
		super_network	Id of super Network entity instance Id
		sub_network	list of Network entity instance Ids
		components	list of entity instance Ids
Comment	<C></C>	id, name	A textual comment, about the associated entity instance.
Shape	<SH>	id, name	Default attributes
		SHAPE	SHAPE=shape of region=[rect | circle | poly | default]
		COORDS	COORDS=coordinates of region.
XML Comment	<!-- TEXT -->		Textual XML comment

As can be seen, all the explicit attributes of each class as well as the implicit ones (-which are part of a class due to the relationships of the class)[3] are sub-entities of block-level[4] entities of the XML entity (of the equivalent class).

Organisational Structure of XML entities for the GTF (excerpt):

Entity (Object)	XML Tag	Sub-Tags	Comment
GTFDB	<GTFDB>		block entity, Toplevel entity enclosing all other entity instances of one data base
		<T></T>	
		<L></L>	

[3] For example: the class Terminator has a relationship „start_of“ / „starts_in“ with Link. This relationship specifies which centroid Node or intersection Node (which are Terminators) is the starting point of a Link. Because of this relationship, each Terminator object (thus each centroid Node or intersection Node object, because these are derived from Terminator) have an attribute pointing to the Link object of which the Terminator is the start of. This specification will refer to these implicit attributes by the name of the relationship, e.g. the relationship „start_of“ / „starts_in“ between Terminator and Link implies an attribute „start_of“ in the Terminator entity and an attribute „starts_in“ in the Link entity.

Inheritance relationships are mapped as sub-tags, e.g. Node is a sub-type of Terminator (Node „inherits“ from Terminator), therefore the Node-tag <N></N> is only allowed as a sub-tag of the Terminator-tag <T></T>.

[4] A block-level entity is an entity that can contain other entities. An inline entity (inline element) is an entity that isn’t allowed to contain other entities.

Entity (Object)	XML Tag	Sub-Tags	Comment
		<F></F>	
		<SP></SP>	
		<V></V>	
		<A></A>	
		<SE></SE>	
		<U></U>	
		<M></M>	
	</GTFDB>		
Terminator	<T>		block entity
		<N></N>	A Terminator can be the parent of an intersection Node
		<C></C>	A Terminator can be the parent of a centroid Node
	</T>		
Link	<L>		block entity
		<LI>	
		<LF>	
		<LC>	
	</L>		
Zone	<Z>		Inline entity
Intersection Node	<N>		block entity
		<NE>	
		<NI>	
	</N>		
SpawnFactor	<F>		block entity
		<FA>	
		<FAR>	
		<FE>	
		<FG>	
		<FP>	

Entity (Object)	**XML Tag**	**Sub-Tags**	**Comment**
		<FL>	
	</F>		
Specification	<SP>		block entity
		<SPM>	
		<SPS>	
		<SPT>	
	</SP>		
Alternative	<A>		Inline entity
Service	<SE>		block entity
		<SEF>	
	</SE>		
Vessel	<V>		block entity
		<VA></VA>	
		<VRO></VRO>	
		<VRA></VRA>	
		<VW></VW>	
	</V>		
Unit	<U>		block entity
		<UDI>	
		<UDA>	
		<UP>	
		<UG>	
	</U>		
Meta	<M></M>		block entity
Network	<NE>		inline entity
Comment	<C>		block entity
	TEXT		Text of Comment
	</C>		

Entity (Object)	XML Tag	Sub-Tags	Comment
Shape	<SH>		Inline entity, all information of this entity are in the entity attributes: id, name SHAPE, COORDS

Example: "Airport Network"

A (simple) Network consisting of an origin Node O, a destination Node D, both linked to an airport Node A:

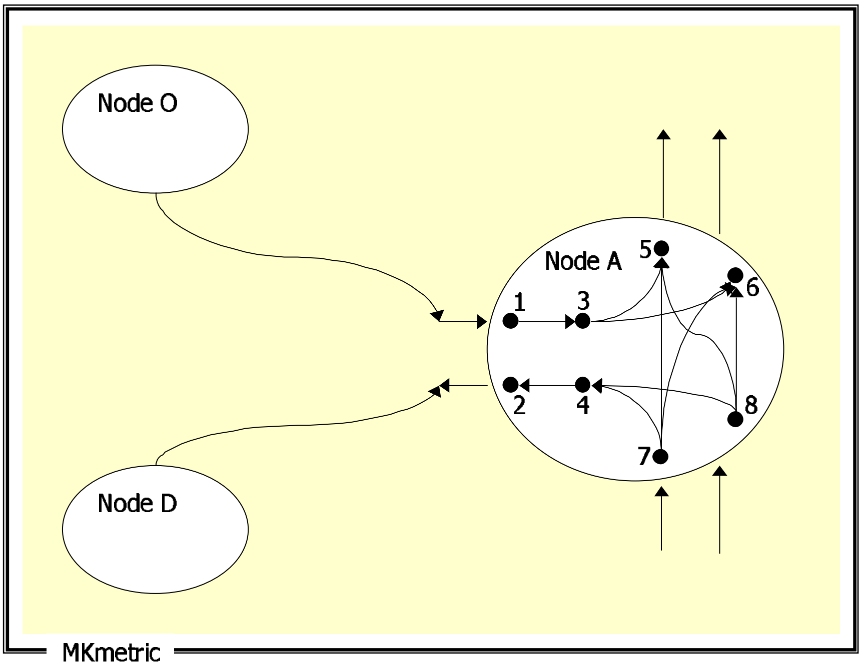


Fig. 14. Airport Node Network

This structure would be represented in a GTF data model using the ad hoc XML format as:


```
<GTFDB id= “1” name= “Airport Network Example”>
<T id= “2” type= ”CentroidNode”>
        <N id= „3” name= “O” type= “1”>
        </N>
</T>
<T id= “4” type= “CentroidNode”>
        <N id= “5” name= “D” type= “1”>
        </N>
</T>
<!-- definition of the internal Nodes -->
<T id= “6” type= “IntersectionNode”>
        <N id= “7” name= “A1 airport access” type= “1”>
        </N>
</T>
<T id= “8” type= “IntersectionNode”>
        <N id= “9” name= “A2 airport egress” type= “1”>
        </N>
</T>
<T id= “10” type= “IntersectionNode”>
        <N id= “11” name= “A3 check-in counter” type= “1”>
        </N>
</T>
<T id= “12” type= “IntersectionNode”>
        <N id= “13” name= “A4 check-out counter” type= “1”>
        </N>
</T>
<T id= “14” type= “IntersectionNode”>
        <N id= “15” name= “A5 departure national” type= “1”>
        </N>
</T>
<T id= “16” type= “IntersectionNode”>
        <N id= “17” name= “A6 departure international” type= “1”>
        </N>
</T>
<T id= “18” type= “IntersectionNode”>
        <N id= “19” name= “A7 arrival national” type= “1”>
        </N>
```

```
</T>
<T id= “20” type= “IntersectionNode”>
        <N id= “21” name= “A8 arrival international” type= “1”>
        </N>
</T>
<!-- definition of the link from Node O to Airport A -->
<L id= “10000” name= “Route 66 to Airport A” type= “LinkInfrastructure” starts_in= “2”
ends_in= “22”>
</L>
<!-- definition of the link from Airport A to Node D -->
<L id= “20000” name= “Highway 928” type= “LinkInfrastructure” starts_in= “22” ends_in=
“4”>
</L>
<!-- definition of the internal links of Node Airport A-->
<L id= “30000” name= “to check-in” starts_in= “6” ends_in= “10”>
</L>
<L id= “30001” name= “from check-out” starts_in= “12” ends_in= “8”>
</L>
<L id= “30010” name= “to departure national” starts_in= “10” ends_in= “14”>
</L>
<L id= “30011” name= “to departure international” starts_in= “10” ends_in= “16”>
</L>
<L id= “30020” name= “from arrival national” starts_in= “18” ends_in= “12”>
</L>
<L id= “30021” name= “from arrival international” starts_in= “20” ends_in= “12”>
</L>
<L id= “30030” name= “from arrival national transfer to departure national” starts_in= “18”
ends_in= “14”>
</L>
<L id= „30031” name= “from arrival national transfer to departure international” starts_in= “18”
ends_in= “16”>
</L>
<L id= “30040” name= “from arrival international transfer to departure national” starts_in= “20”
ends_in= “14”>
</L>
<L id= “30041” name= “from arrival international transfer to departure international” starts_in=
“20” ends_in= “16”>
</L>
<!-- definition of the Airport Node-Network-->
```

```
<T id= "22" name= "Airport A">
            <N id= "23" type= "5">
                  <NE id= "24" name= "Airport A" components=
"6,8,10,12,14,16,18,20,30000,30001,30010,30011,30020,30021,30030,30031,30040,30041"></
NE>
            </N>
</T>
</GTFDB>
```

2.5.3 Transportation–data Interchange Protocol (TIP)

Section 2.5.2 "Generalised Transportation-data Format (GTF)" established the specification for a GTF Data Model. Now a specification concerning the available commands to a user's workspace is required. These commands will be part of a GTF file and will enable a model provider to process the GTF data file so that the requested answers are computed. This is necessary, because a GTF Data Model alone doesn't contain any information on what shall be done with the data. This is where TIP is necessary. TIP is a generalisation of "usual" commands (queries) to a transportation model. The development of TIP is based on the classic four step transportation model: generation, distribution, modal split and assignment. Within these four stages, a number of commands (independent of the actual model or the model's philosophy) can be issued to the model in order to produce intermediate data or final model results. These results can then be passed through a filter defined in a TIP command file that is part of a GTF-XML file. The filter extracts the data relevant to the user's query out of the model results and notifies the user's system that the requested results are available for download from the model provider.

2.5.3.1 Classification of possible queries

The categories of possible (transportation–)information exchange are:

- pricing policies
- regulatory policies
- investment policies
- co–operation of models

The following types (per category) are feasible – also ensuring that meaningful results could be produced:

modification of model input

1. input modification, e.g.

proportional modification of a variable's value on a whole network or a specific sub–set (for pricing policies, regulatory policies)

modifications of networks (for investment policies)

2. output queries, e.g.

modal split effects (e.g. high speed train vs. air; alternative i vs. alternative j)

generation and distribution effects (e.g. on airport choice results)

communication between models

1. Model 1→ Model 2: output of Model 1 (e.g. passenger movements or OD–flow matrix) as input to Model 2

2. Model 2 → Model 1: output of Model 2 (e.g. modal split matrix) Model 2 as input to Model 1

With respect to scenario definitions and future projections the described options fit into the following framework (please refer to Fig. 2. Queries & Models). For each of the components in the last level of the hierarchy two commands must be available.

Input modification

1. explicit change of variable values, e.g. variable X = 100

2. functional change of variable values, e.g. variable X = (variable Y * 2) + variable Z (all mathematical standard operators and functions are allowed for manipulation, e.g. log(), sin(), +,–,*,/, exp() etc.)

Output query

3. output matrix to be calculated, e.g. modal split for all available modes, assigned road network – Germany

4. definition of extracted variables, e.g. modal split of mode road and air, travel–time on link 1152, travel–time of shortest path between zone 51 and zone 894

The variables available in 1., 2. and 4. are the attributes of entity instances defined in the GTF data model.

2.5.3.2 TIP commands

The commands needed are split into two categories 1. manipulation of variables (selecting & setting / updating) 2. creating, requesting matrices (selecting & calculating). These categories are based on the usual structure of transport models, please refer to Fig. 1. Policy Scenarios & Queries.

1. selecting & setting / updating

The variables available for manipulation are those defined in the GTF Data Model, e.g. "entity SpawnFactor – Population Class 79 – INSTANCE 34923" for a *single* manipulation or "entity SpawnFactor – Population Class 79" for manipulation of *all* instances. The semantics for the manipulation commands is based on SQL, because the manipulation of GTF variables is a manipulation of relational data. The manipulation commands (i.e. manipulation of model input data) always refer to data already located at the model provider[5]. The commands have the following syntax:

```
UPDATE <entity>.<ALL|SINGLE> SET <variable>=<value>
UPDATE <entity>.<ALL|SINGLE> SET <variable>=<function>
UPDATE <matrix>
```

Where “function” is a mathematical function of any variables in the GTF data already at the model provider. A user can either modify single data elements (e.g. travel–time on link between node 42873 and node 42192 multiplied by 1.2) or lists of data elements (e.g. all the travel–times in a network multiplied by 1.1). <matrix> is one of those specified in the paragraph below.

2. selecting & calculating

The requests for calculation are related to the usual phases that a transportation model comprises: generation (production / attraction), distribution, mode choice (modal split), traffic conversion, route choice and assignment (see Fig. 2. Queries & Models). The request commands (i.e. model output data) are introduced by the keyword ”CREATE”

Command syntax: “CREATE <matrix> {MODE|PURPOSE|SEGMENT|PRODUCT}”.

Followed by the these keywords

KEYWORD	**Matrix contents**
GENERATION	Matrix: ZONE (x PURPOSE or SEGMENT / PRODUCT) Cell contents: ZONE number of TRIPS or amount of FREIGHT
PRODUCTION / ATTRACTION	Matrix: ZONE x ZONE (x PURPOSE or SEGMENT / PRODUCT) Cell contents: number of TRIPS or amount of FREIGHT
DISTRIBUTION	Matrix: ZONE x ZONE (x PURPOSE or SEGMENT / PRODUCT) Cell contents: number of TRIPS or amount of FREIGHT
MODAL SPLIT	Matrix: ZONE x ZONE x MODE (x PURPOSE or SEGMENT / PRODUCT) Cell contents: amount of FLOW / PERCENTAGE (trips / tons)

[5] To manipulate data located in the GTF file doesn’t make sense, as the result of the manipulation can be computed beforehand, at the user’s site.

TRAFFIC CONVERSION	Matrix: ZONE x ZONE x MODE (x PURPOSE or SEGMENT / PRODUCT) Cell contents: number of VEHICLES
ASSIGNMENT	(loaded) network: NODE x NODE Cell contents: LINK attribute(s)

The keyword defines which output matrix shall be computed and transmitted (after filtering) back to the user. The specification of MODE, PURPOSE, SEGMENT / PRODUCT is optional. If one is specified it must follow the keywords preceding.
For example: “CREATE DISTRIBUTION BUSINESS”.
The output filter is defined with

```
FILTER <matrix> <variable 1> ... <variable N>
```

The meaning of this line is: ”Filter from the output matrix <matrix> the variables <variable 1> through <variable N>”. Where <variable> is the fully–qualified[6] name of an entity attribute.
TIP commands are defined in the XML segment <TIP></TIP>.

2.5.4 IMPLICATIONS / RAMIFICATIONS OF GTF

The impact of GTF software might have many ensuing commercial and practical ramifications:

1. people dealing with problems appearing in different working areas can exchange information, e.g. analysing side effects when changing from a higher to a lower aggregation level
2. synergetic effects can result from the possibility of transferring knowledge between systems and points of view
3. it will be possible to compare different models’ results (and their quality) as the models can be used on the same data(–base)
4. model users won’t always have to (re–)create their own databases over and over again like in the past, but will have access to standard data(–bases)
5. data–(bases) will gain in quality as time passes, because the data providers will have an incentive to update their databases regularly and properly, since only the “good” databases will be used
6. the fast pace at which telecommunication and telematics are developing in respect to the pace at which models are developing, suggests that within a few years users will be able to use even remote models interactively

[6] A fully-qualified name consists of the name of the entity preceded by the names of all parent entities.

7. researchers from different countries can work on the same database and exchange knowledge about their findings
8. users will request new models or combination of models, which previously could have been denied by the consultants, because of lack of transparency on the supply–side of the business
9. the clients / users will have the possibility of choosing and combining models by choosing model outputs to be fed as input to other models to compute further results. This might even be done independently and automatically. Like this the client will not be bound to any specific model but will have the freedom to choose the “best” models for the task / question at hand.
10. all these effects will have a vigorous impact on research in the modelling and other fields

2.5.5 SUMMARY

GTF is an acronym for “Generalised Transportation-data Format” specification. The goal of GTF is to standardise the information used by transportation modelling software for the purpose of electronic data interchange (EDI). The GTF specification uses already defined standards wherever possible in order to maximise acceptance and to minimise redundant work.
To accomplish this the GTF specification comprises the following parts <u>1. a standardised definition of transport information</u>, but without constraining the possible information to any specific sub-set. This is called the “GTF data model”. See section 2.5.2 “Generalised Transportation-data Format (GTF)”. <u>2. a standardised set of commands to run models</u> and to retrieve relevant data. This is called TIP (“Transportation-data Interchange Protocol”). See section 2.5.3 “Transportation–data Interchange Protocol (TIP)” <u>3. a standard format for arranging data in a file used for EDI</u> and a standard protocol for exchanging the data file. For this XML is used. See section 2.5.2.5 “GTF data model & GTF–XML message specification”. The goal that was defined for GTF was “GTF is for the exchange of information between models and other models & softwares.” it does not entail “... exchange of model databases.” The difference is crucial as the first definition doesn't require the GTF to be in anyway optimised for database handling. GTF is only for a definition of a data model that can be used for a specific database implementation. In contrast, the second definition implies this optimality. What must be clear is that taking a GTF file and putting it into a database such that the access of the data in the database is optimal (i.e. the table definitions are such that, e.g. a minimum of select statements are needed) must be

done by a “GTF Translator”. This means that providers adopting GTF need to define 1. their optimal database configuration for their needs 2. to implement a translator that takes the GTF file and converts it into the optimal database tables. Since different people / users have different optimality criteria, GTF defined in this paper is not for optimal database use, but for completeness of the description of the transportation problem domain.

In GTF, the term “data” is simply a fact like “the number 50”, while the term “information” is the association of “data” with a meaning, e.g. “the number 50” and “speed in km/h” which together give “a speed of 50 km/h”. This kind of association between data and meaning is needed to specify exactly what the data signifies. For example “the number 50” when associated to “speed in miles / h” has a completely different sense compared to “speed in km / h”. The former means that “the number 50” is approx. 1,6 times larger than in the latter case. In general, “information” is data with a meaning. By associating data with meaning, the more general “information” can be described. “Data” is fact without exact meaning. Transport models are very sensitive to the information used as input. For example, if a model runs using the information “speed is measured in km/h”, the same model most probably won’t produce valid results, if it is fed with data based on the information “speed is measured in miles / h”. For sophisticated models, it is not enough just to send “data” when exchanging information between models, but the exact meaning must be transmitted, too. To accomplish this, GTF specifies a generalised framework of information in the context of transportation. Within this framework a set of defined pieces of information can be taken and arranged in order to convey the proper meaning for the data being exchanged in EDI, somewhat like building something with LEGO. These pieces of information are contained in “entities” (class, type of information) and “relationships” defined in a so called “data model” by the “GTF specification” which defines the “GTF Data Model”. In order to convey the meaning of a bit of data, different entities can be linked through predefined relationships. The entities and relationships used are then transmitted in an EDI together with the data. GTF is not an imposition of a format, but is a flexible way of describing transportation information.
Once the information is transmitted to a model provider one wants to use the model on the information and one wants to retrieve the results. For this purpose, GTF specifies a set of commands. These aren’t related to the workings of any particular model, but are related to the retrieval of information results once the model has computed the input. In

the transportation context, models produce results at some stage or the other during the computation. These results can be classified in a standard way using usual requests by clients of model providers. These requests are for example: a flow matrix, modal split, shifts in service levels etc. GTF specifies a set of standard commands relating to such requests, which can be satisfied by most (any) model(s). The set of commands is called TIP (Transportation-data Interchange Protocol). The term “GTF specification” implies both “the GTF data model” and “TIP”.

To be able to exchange data electronically, the data must be in a form that can be processed automatically. This is done by specifying the arrangement of the data within an EDI file and the protocol to interpret the sections in an EDI file. Furthermore, to ensure maximum portability across very different hardware and software platforms (e.g. the sender uses UNIX/Solaris and the receiver uses PC/Windows) the transmission files must be in ASCII. This is resolved by using XML and defining a GTF-XML.

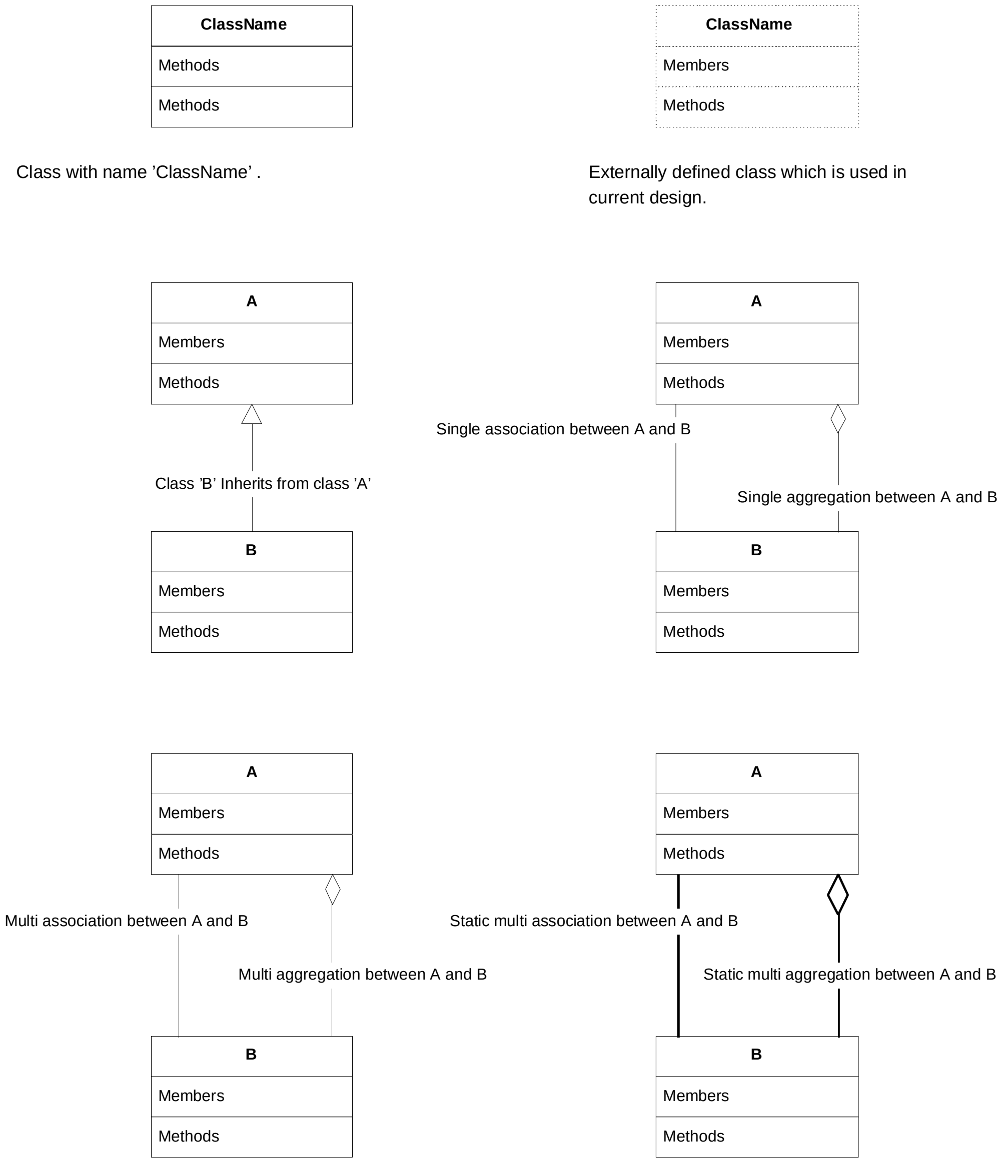
The structure of GTF is open and by following the defined rules it can be enriched, detailed and extended in nearly any direction concerning transport. This specification doesn’t cover everything in detail, but tests showed that models of urban transport, freight and passenger models, special models for shipping, road specific information on load or damages, schedules as well as indicators or indexes can be handled by GTF. It’s a matter of effort to enrich the initial specification developed in [MKMETRIC99].

In the wider context, GTF is for linking (mainly) passenger transportation models to any system. The GTF specification was developed to enable model providers to offer their transportation models’ results in a standard fashion. Subsequently, this enables computer systems to present the results in the form a user wishes. Most commonly, users want to view the results using standard applications, e.g. Spreadsheets, Desktop Mapping etc. Therefore, computer systems should also offer links between GTF translators and standard applications that are previously registered in the system. A complete system furthermore should assist the user with the tasks of finding appropriate data and appropriate model providers to answer a user’s transportation query. Fig. 13. Typical Exchange Situation shows the typical environment computer structure where GTF is to be used.

2.5.6 DEFINITIONS, ACRONYMS, ABBREVIATIONS AND USED SYMBOLS

Acronyms

GTF	Generalised Transportation-data Format
TIP	Transportation-data Interchange Protocol
EDI	Electronic Data Interchange



UML	Unified Modelling Language
OO	Object-Oriented
ITP	Intraplan
NUTS	Nomenclature of territorial units for statistics

Fig. 15. Explanation of the symbols used

2.5.7 REFERENCES

Literature

Brown, David, “An Introduction to Object-Oriented Analysis: Objects in Plane Language”, John Wiley & Sons, 1997
Budd, Timothy, “An Introduction to Object-Oriented Programming”, Addison-Wesley, 1997
Button, Kenneth J., “Transport Economics”, Edward Elgar Publishing Limited, 1993

Chen, P.P., “The Entity–Relationship Model — Toward a Unified View of Data”, ACM Trans. on Database Systems, Vol. 1, No. 1, March 1976, pp. 9-36

European Commission - Directorate General Transport, “Transport Research APAS Road Transport VII - 34, Network Architecture”, 1996

EUROSTAT, “GESMES 93 –Exchange of Multidimensional Statistical Arrays and Time–series Data. Volume 1: Guidance to Users, Volume 2: Reference Guide”, EUROSTAT, Luxembourg, 1996

EUROSTAT, “Glossary of Transport Statistics”, EUROSTAT, Luxembourg, 1994

EUROSTAT, NUTS. EUROSTAT, Luxembourg, 1995

Gamma, Erich et al. (Gang-Of-Four), “Design Patterns”, Addison-Wesley, 1994

Mandel B., Ruffert E., “GTF Final Report”, MKmetric GmbH, 1999

Mandel B., Ruffert E.; “Generalised Transportation-data Format (GTF): Data, Model and Machine Interaction” paper presented at the 1st ITEM Workshop; Montréal/Canada; 2000

Mandel B., Ruffert E., “Steps towards the infrastructure of ETIS (European Transport Information System) from a user’s point of view – The way towards an operational ETIS using synergies between 4th FP outcomes”; Mesudemo Workshop 2; Rotterdam; 06/1999, MKmetric GmbH, 1999

Martin, Rhiele, Buschmann “Pattern Languages of Program Design”, Addison Wesley, 1998

Marchal, Benoit, “XML by Example”, Que; ISBN: 0789722429, 1998

Moellering H & Hogan R “Spatial database transfer standards 2”, Elsevier Science Ltd., Oxford., 1997

Ortúzar, J. de, Willumsen, L.G., “Modelling Transport”, John Wiley and Sons, 1990

NACE, http://www.cdnet.at/internetpages/cgi/webc.exe/german/nace.htm, ISO 3166 Maintenance Agency

National Institute of Standards and Technology, “NIST Integration Definition for Information Modelling”, Federal Information Processing Standards Publication 184, NIST Gaithersburg, MD., 1993

Rumbaugh J, Blaha M, Premerlani W, Eddy F & Lorensen W “Object–Oriented Modelling and Design”, Prentice Hall, New Jersey, 1991

UML, resource (documentation etc.) http://www.rational.com/uml/index.jtmpl

Projects

“BRIDGES”, “Building Bridges between Digital Transport Databases, GIS Applications and Transport Models to Develop ETIS Software Structure” (contract no. ST-96-AM-1138), on behalf of the Commission of the European Community – DG VII, 1997-1999

“Spotlights(TN)”; “Scientific forum for making advanced transport models fully transparent to end-users, open and more integrated into policy-making”; on behalf of the Commission of the European Communities– DG-Energy and Transport; Actual Cost Contract No.: 1999-TN.10941; 2000-2003

2.6 MESUDEMO WORKSHOP 2 – ROTTERDAM – 17TH-18TH OF JUNE 1999

MESUDEMO WORKSHOP 2 – ROTTERDAM – 17TH-18TH OF JUNE 1999

Steps towards the infrastructure of ETIS from a user's point of view

- The way towards an operational ETIS using synergies between 4th FP outcomes -

Benedikt N. Mandel and Eduard Ruffert

MKmetric Gesellschaft für Systemplanung mbH,
Durlacher Allee 49,
76131 Karlsruhe, FRG
email: *mandel@mkm.de* and email: *ruffert@mkm.de*

Abstract The goal of ETIS is to support policy-makers during the whole process of "Policy Scenario" formulation through to making the actual decision. Outcomes from the BRIDGES project are usable for the implementation of ETIS. The outcomes from BRIDGES can be used to guarantee homogeneous data, data structures, information contained in the data, software and software structures. The main concepts are a number of guides which serve as directories of available data (Digital Data-sources Guide - DDG), models (Digital Models Guide - DMG) and compatibility between both and between the models (Digital Models / data Compatibility Guide - DMCG), software to glue applications together that weren't initially developed to communicate with each other (Network Information System – NIS), a homogenous data (information) exchange format (Generalised Transportation-data Format – GTF and GIS-GTF) and a user-friendly interface (Decision Support System – DSS). Together, these components can be composed to be the foundation of ETIS. With further development, of a generic language for accessing models and retrieving results (Transportation-data Interchange Protocol – TIP), DMG, DMCG, further enhancement of the DDG, GTF, DSS and development of a single access point to all the functionality through a Web-Interface (i.e. Web-page / site for the Internet) an operational ETIS can be implemented. The ETIS structures were conceived to be "self-cultivating", i.e. only minimal effort from outside (e.g. the European Community) will be needed to maintain ETIS and for ETIS to prosper and grow. The "ETIS club" will make sure of this without having to explicitly make the effort, because of the "invisible hand" of fair regulations, quality controls and market competition.

Keywords ETIS, club, BRIDGES, implementation, models, software, data, homogenous, communicate, accessing, retrieving, source, guide, compatibility, decision, European, Commission, EC, GTF, TIP, DDG, DMG, DMCG, NIS, DSS.

Table of contents

Table of figures

2.6.1 CURRENT SITUATION & PROBLEMS

This chapter briefly describes the well known interdependency between the elements of policy scenarios a decision maker wants to analyse. Queries reflect the user's request to a model and finding the data pool which has to be accessed. They also identify the basic problem decision makers are faced when they want to validate different analyses leading into the problem of incompatibility of results.

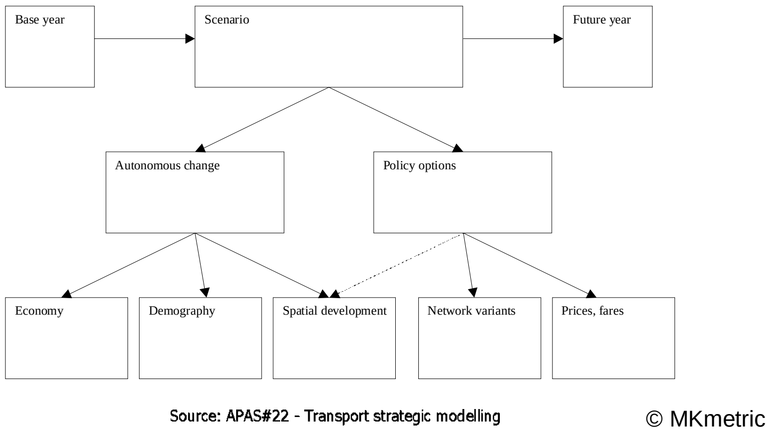


Fig. 16. *Policy Scenario & Queries according to APAS*[7]

The goal of ETIS is to support policy-makers during the whole process of "Policy Scenario" formulation through to making the actual decision. Fig. 16 depicts the usual inputs that define "Policy Scenarios", like economic, demographic and spatial developments as well as network changes and changes in prices and fares for the use of transport supply. This is the level at which decision-makers operate therefor this should be the domain from which the User Interface (UI) of ETIS is derived from. In this way, the users (decision-makers) won't need to think in different categories other than the ones that are known and common to the user. This is a major problem nowadays, because the systems to support the decision-maker require that the decision-maker thinks in different categories other than the ones he is used to.

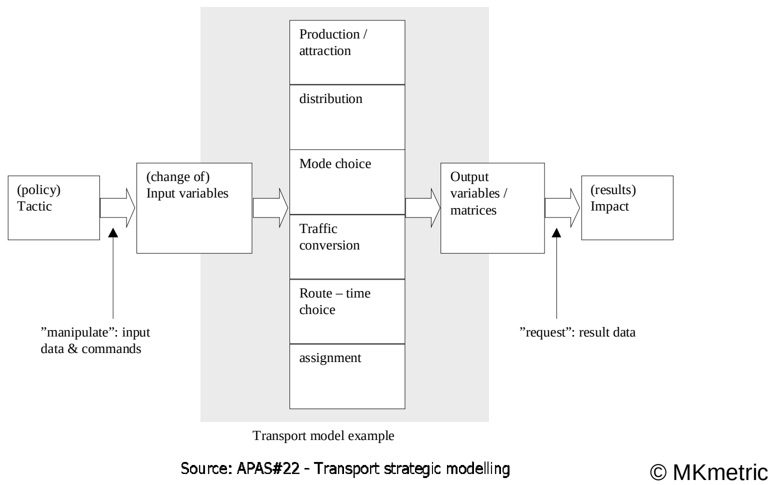


Fig. 17. Queries & Models

[7] "Transport strategic modelling", Transport Research - APAS, 22, European Commission – Directorate General VII - Transport

The usual workflow currently, is depicted in *Fig. 17*. The policy to be analysed using models is “translated” into model specific commands such that the concrete model, that is chosen by the user “by hand” and without support, is run and produces results – that are “hopefully” the results required for the user to make an objective decision for the problem at hand.

Despite the initial problem to formulate policy scenarios, it is even harder to translate them into queries which can be understood by the addressed models. This is mainly due to fact that different models need different queries because of the inhomogeneous structure of the underlying data model and theory of the used model. Furthermore there is the problem to link the data with the models in a way that is usable by the models. Currently, the structure of the numerous software applications and databases is inhomogeneous and largely incompatible with each other. Which, very frequently, leads to the problem (- rather the impossibility) of comparing results from scenarios based on different software applications and databases.

The basic problem of trying to compare results from different software applications and databases lies in the fact that the underlying data-models (DM) for the software applications and databases are incompatible or only made compatible with extreme difficulty and loss of information. This is because the informational units used by the data-models are to “large”, meaning one unit contains too much information.

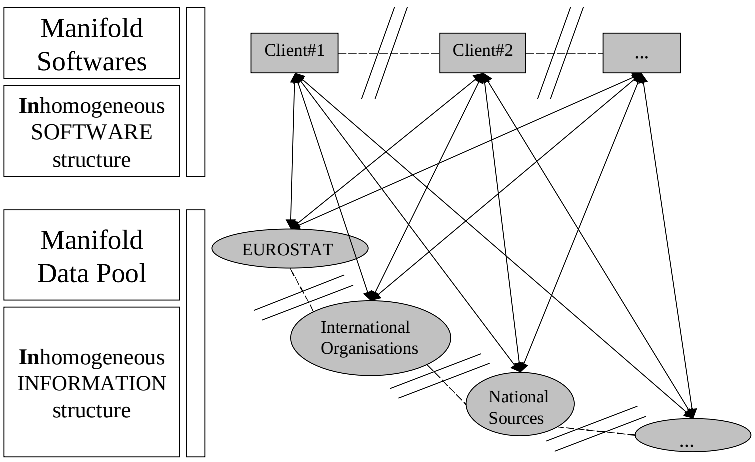


Fig. 18. Current Data Pool

Therefor it is virtually impossible to manage the task of matching informational units contained in different databases without loss of information.

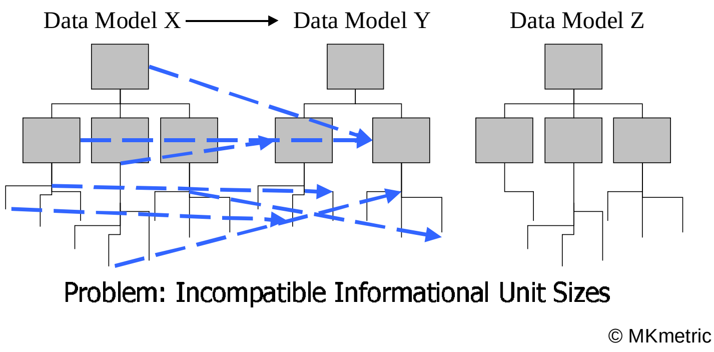


Fig. 19. Matching Data Models Step 1

If more than one matching of informational units in different databases is required, there is a loss of information, in each step of transformation. Which very often (almost in all cases) leads to the fact, that the output of all the transformations won't contain all expected relevant information.

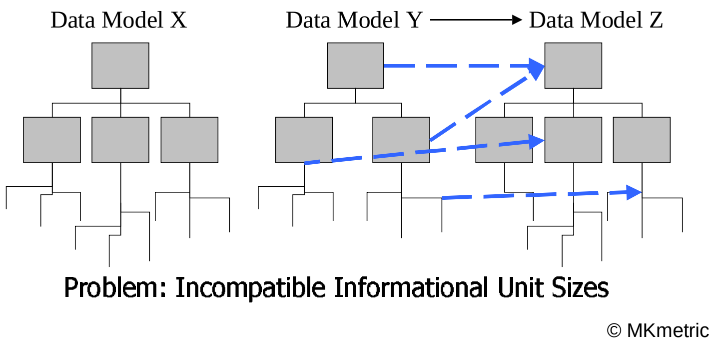


Fig. 20. Matching Data Models Step 2

Therefor decision makers are always faced with the problem of the incompatibility of results when taking a decision.

2.6.2 ETIS VISION & CONCEPT

Starting from the ETIS vision where any user (all over the EC) can access data, models, analyses and GIS tools within his workspace through a user-friendly and guiding interface, the discussion about the necessary concrete developments will be done step by step laying out an adequate and ideal ETIS concept.

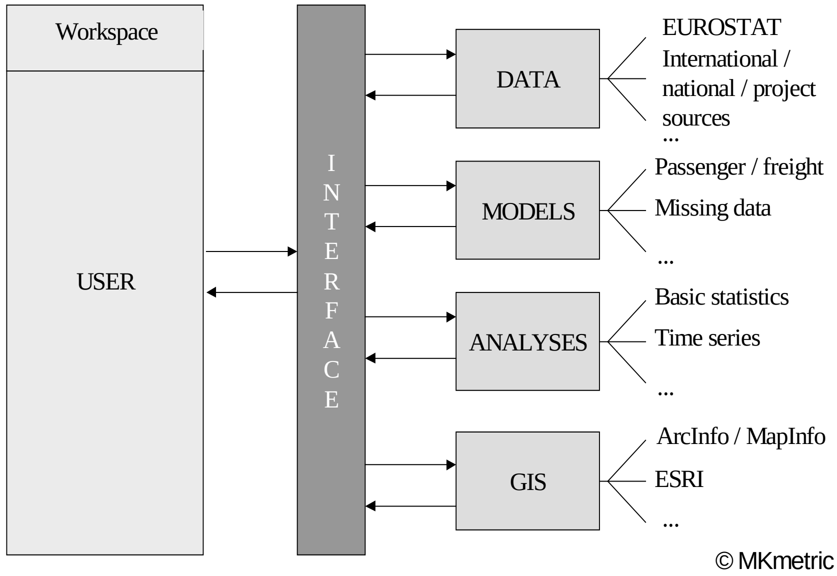


Fig. 21. ETIS Vision

So the first step is to increase the technical accessibility to data in an automatic manner as stated in chapter 1. The aim is to keep the plurality models, data and all other tools but to ensure that information can be exchanged without any loss of content. Within ETIS the structure of numerous software applications and databases have to be accessible in a homogeneous and compatible fashion based on an uniquely identifiable "Generalised Transportation-data Format" (GTF). A set of GTF translators will ensure that full information exchange is possible and provide a single access point to all models, data, analyses and GIS tools.

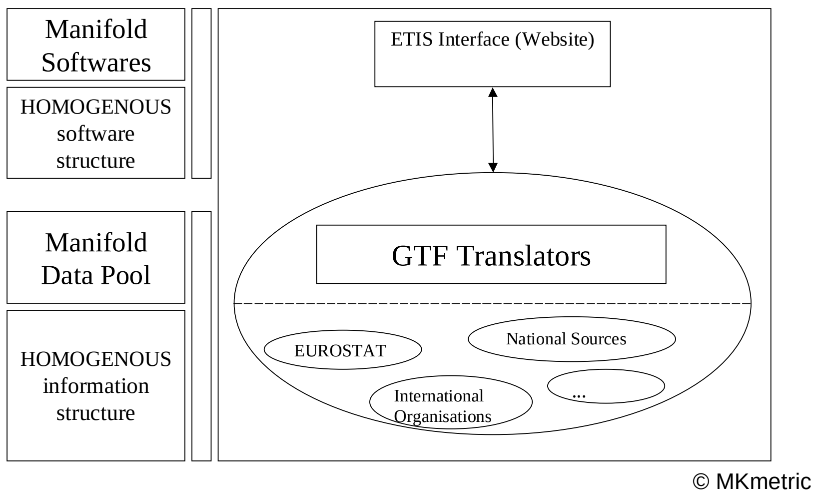


Fig. 22. ETIS / GTF Data Pool

Following the ETIS vision and considering all the problems described in chapter 1, the structure of ETIS should be the following:

Instead of having disparate and manifold software applications and databases, ETIS should contain all necessary elements and provide one single access point – through its interface

Instead of having incompatible proprietary formats and informational contents, a “Generalised Transportation-data Format” (GTF) should be used throughout the whole system, by providing translators to / from the proprietary formats to GTF. GTF consists of a generalised data-model (GTF-DM), a standard exchange format (GTF-GESMES) and a generic language for accessing models and retrieving results (“Transportation-data Interchange Protocol” – TIP)

The interface of ETIS should provide a support system to aid the user in formulating policy scenarios in the way that the system can decide which database(s) and model(s) should be used to compute the results or if the users wish to take these choices themselves, to aide the users to find the relevant data sources and model providers – this is a “Decision Support System” = DSS.

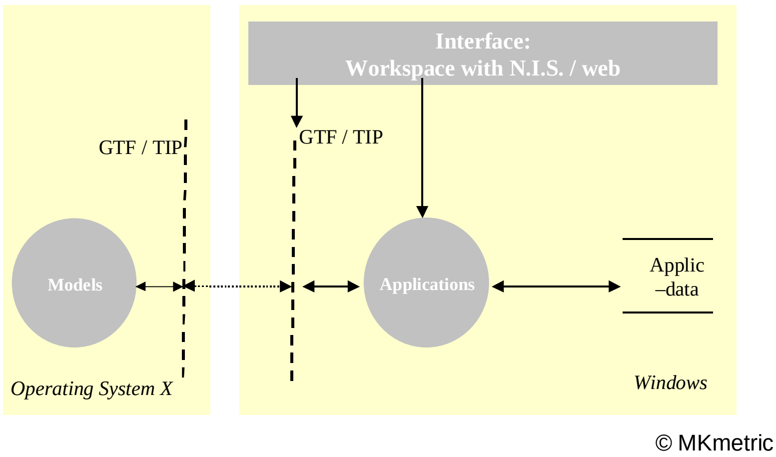


Fig. 23. ETIS Structure Overview

And lastly, everybody and everything that is involved in ETIS is called the “ETIS club”.

In a bit more detail, the required ETIS structure has the main components and the following infrastructure:

a software to enable communication between applications that aren’t initially able to communicate with each other (NIS – Network Information System)

a DSS (Decision Support System) to assist the user in formulating and running the queries to solve the policy questions

a DDG (Digital Data sources Guide), a DMG (Digital Model Guide) and a DMCG (Digital Model/data Compatibility Guide) are needed for the DSS operate

a TIP (Transportation-data Interchange Protocol) to run models using a “generic language”

a GTF translator for each model that is to be accessed by the system

and relevant databases in GTF for input to the models

A filter should be used to make sure only absolutely necessary data will be transmitted to model providers (- in order to decrease the necessary bandwidth). Also, it should be required to transfer either only incremental data or only complete data, depending on which one has the lesser amount of data. *Fig. 24* depicts these concepts.

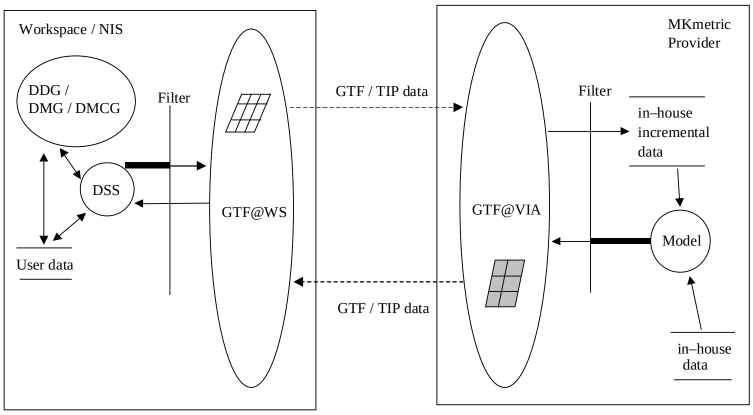


Fig. 24. ETIS Concept

As many ETIS structure details are mentioned in the ETIS concept, the following chapters will shed a light on these “infra”-structure elements.

2.6.3 INFRASTRUCTURE FOR ETIS

In this chapter we want to explain the basic infrastructure details necessary for the ETIS concept depicted in *Fig. 24*. Therefor you’ll find the principles of how the communication, automation, user interface and guides support ETIS, in this chapter. Some of the supporting features are already worked in out in quite some detail within the BRIDGES[8] research project of the 4th FP of the EC. It should also be stated that these infrastructure elements have been (and sill are) implemented and tested at the EC-DGVII, EUROSTAT and European Investment Bank (EIB).

2.6.3.1 Supporting Model Communication

GTF - the “Generalised Transportation-data Format” – is a data-model and exchange format specification. Software applications that implement this specification for a specific model are called “GTF Translators”.

The data-model underlying GTF is generic and atomic. This means, that the information units in the GTF-DM are small – reduced to the absolute minimum of one bit of information. This DM was specifically developed to support the informational contents of as many models as possible, i.e. the units are “atomic”. Therefor the DM was based on the information that is needed in the exchange between models and not based on any DM of any specific model(s), i.e. GTF is generic.

With GTF, a homogeneous view on a specific data-model of a transport model can be defined. Thus, GTF can be used as mediator format while transforming one data-model of the output of a model to another.

GTF thus also reduces the amount of required translators from n*(n-1) to n (for each translation direction).

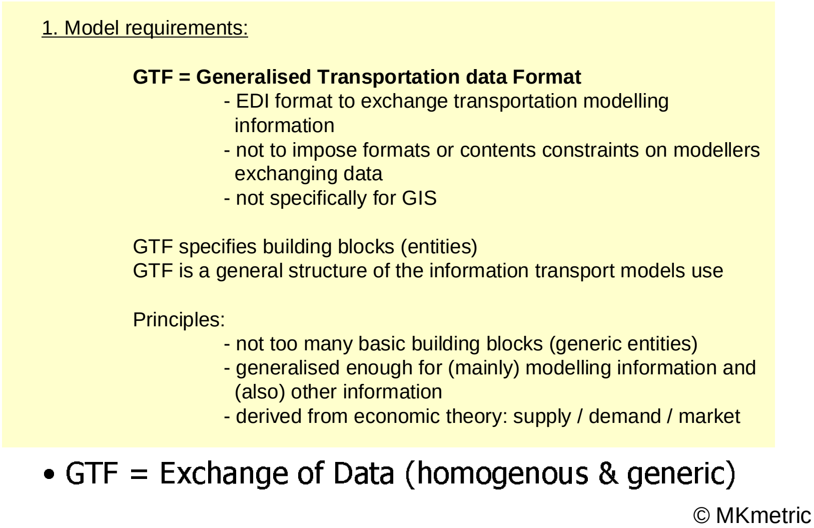


Fig. 25. Supporting Model Communication 1

Subsequently, GTF can be used for exchanging data between models.

1. Model requirement:

GTF = Generalised Transportation data Format

GTF is an EDI format to exchange transportation modelling information

a. between transportation models

b. between transportation models & other applications

GTF is NOT to impose formats or contents constraints on modellers exchanging data

GTF specifies building blocks (entities) which can be used to describe & exchange data used by transportation models.

I.e. GTF gives a general structure of the information transport models use in order to be able to exchange this information.

Principles guiding the design of the GTF data model:

- not too many basic building blocks (generic entities)

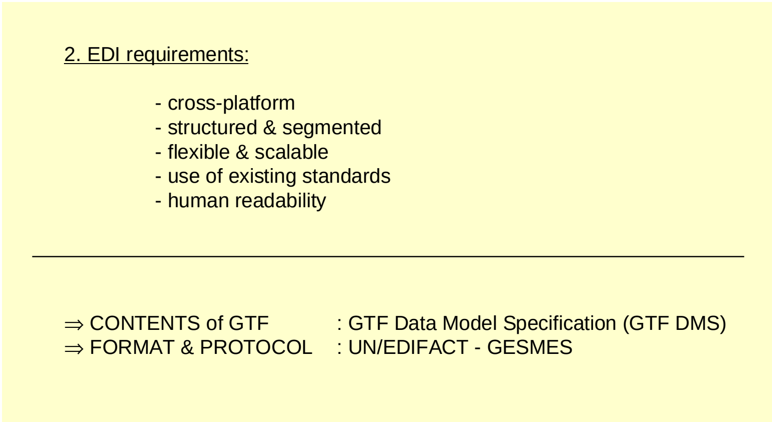
BUT the possibility to cast complex information into such structures (or combinations of these)

- generalised enough to encompass (**mainly**) modelling information, e.g. socio-economic data, (and **also** encompass other information like simple graphical data or GIS information).

[8] "BRIDGES – Building Bridges between Digital Transport Databases, GIS applications and Transport Models, to develop ETIS software structure", 4th FP

- derived from **economic theory: supply / demand / market**

because of the nature of the information to be exchanged. I.e. transportation information like socio-economic data is very much correlated with economic concepts and terminology



© MKmetric

Fig. 26. Supporting Model Communication 2

2. EDI requirements due to ETIS structure and GTF:

- cross-platform = PC - Windows / UNIX -> ASCII
- structured & segmented = for minimal redundancy and reduction of complexity

=> divide & conquer

- flexible & scalable = to adapt to (user / workspace) requirements
- use of standards = so that people use it
- human readability = if problems arise it is always a better thing to have data

readable by humans, directly

=> ASCII

⇒ CONTENTS of GTF is specified in the GTF data model which gives the structure of the entities

⇒ FORMAT & PROTOCOL for transmitting GTF files : UN/EDIFACT - GESMES specification STANDARD

2.6.3.1.1 Matching Data Models using GTF

As described before, the informational units of the GTF-DM are “atomic”. Therefor the informational units (- the data) of any other DM (DM-X) can be decomposed according to the GTF-DM.

For example, an information unit, say “Time” is “50”, of some data-model DM-X of “Flows”, is a non atomic information because one can’t know whether there is much implicit information or not, without having the exact definition of the data-model (and thus the exact definition of the attribute “Time”). “Time” might be an aggregated value meaning the “Total travel time”. This, too, is aggregated and could mean f.i. the sum of access / egress, taxiing time, flight time, etc. Already this example shows that implicit information can be numerous.

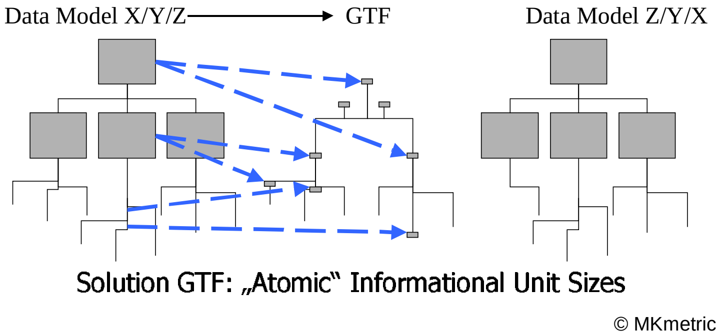


Fig. 27. Matching Data Models using GTF Step 1

To avoid this confusion, the GTF data-model offers containers of “atomic” information which must be used to describe aggregated information as in the example above.
In the GTF-DM, all pieces of information that qualify a piece of data are kept in separate entity instances which are linked through relationships to the entity instance containing the piece of raw data. In the case of the example, this means, that “access/egress”, “flight time”, “total time” etc. are qualifiers (“flags”) attached to the raw data “50”.

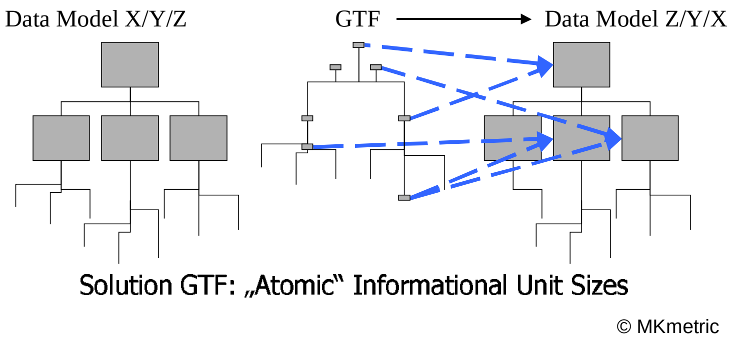


Fig. 28. Matching Data Models using GTF Step 2

Following the example in the previous *Fig. 27*, the decomposed information (DM-X) can be built together differently than the original DM using the GTF informational units, creating the data according to a different DM (DM-Z).
I.e. the assembling of aggregated information units from the GTF-DM units can be done in analogy of decomposing aggregated information units into GTF-DM units. It’s just the reverse principle.
The next sections will describe in more detail the GTF-DM.

2.6.3.1.2 GTF Data Model Overview

The basic information captured in the GTF-DM are:

- the infrastructure elements of networks: NODE, INFRASTRUCTURE-LINK etc.

- the zonal elements needed for flow assignment: ZONE, FLOW and CONNECTOR-LINK

These types of basic information are further sub-divided in the GTF-DM until the level of detail required (one bit) is reached. Also, all entities of the GTF-DM have a “GIS” part and a “TYPE” part. The “GIS”-part is used to capture graphical information, e.g. co-ordinates, vertices attached to LINKs etc. The “TYPE” part contains the information relevant to models, e.g. INFRASTRUCTURE-LINK and FLOW-LINK.

Many usual elements found in the inputs used by models, e.g. node, link, are generalised in the GTF-DM, e.g. a LINK is a connection between two TERMINATORS, a TERMINATOR is a starting point or ending point of a LINK, a LINK can be an INFRASTRUCTURE-LINK, a FLOW-LINK or a CONNECTOR-LINK etc., to provide atomic informational units.

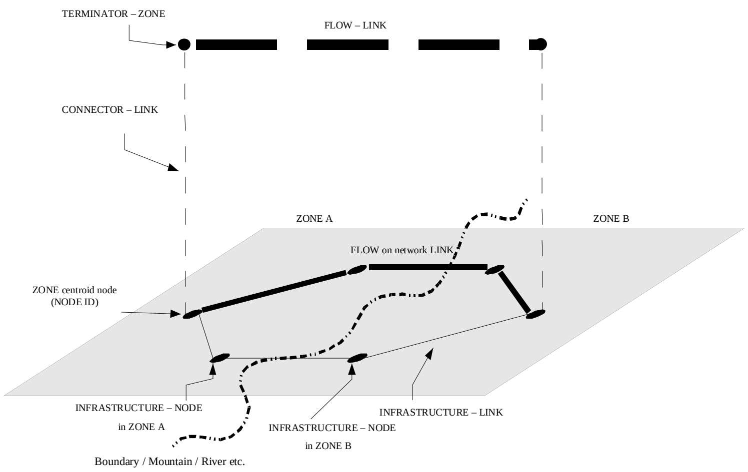


Fig. 29. GTF Data Model Overview

A brief overview covering the GTF entities and relationships is depicted in *Fig. 30*. For further information concerning details we refer to deliverable D3 of the BRIDGES project.

The overview shows the main entities and relationships in the GTF-DM.

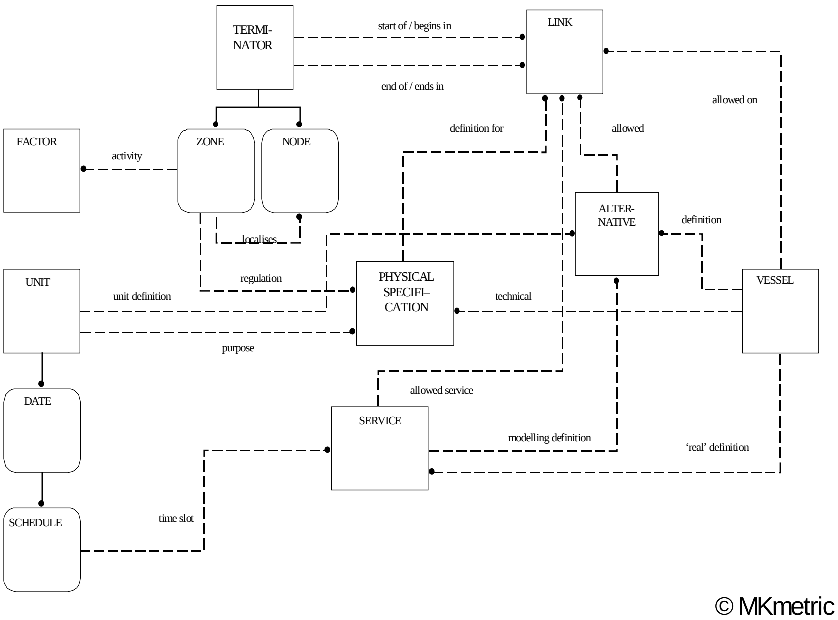


Fig. 30. GTF Data Model Entity-Relationship Overview

To explain the principles of the data model structures as an example, the informational unit "NODE" structure as depicted in *Fig. 31*:

all NODEs have "TYPE" information and (optionally)"GIS" information associated to them
The "GIS" information of a NODE are, e.g.:

"SYMBOL": definition of the symbol to use when displaying the NODE. This entity has a relationship to a "SHAPE" entity which captures the co-ordinates defining the "SHAPE".

The "TYPE" information of a NODE are:

"INFRASTRUCTURE", defining the NODE as a network infrastructure node without further structure

"(sub-) NETWORK", defining the NODE as a sub-network, which itself is a grouping of (possibly) all other entities in the GTF-DM. This means, different levels of detail are possible for a network.

All NODEs have a relationship to a ZONE (- the name of the relationship is "is in"), defining the ZONE in which a NODE is contained. (All ZONEs have relationships to NODEs (- the name of the relationship is "localises") defining the connector NODEs which connect the a ZONE to the network).

A NODE is a child of TERMINATOR (like ZONE). But the function of NODEs is to represent (in combination with LINKs) the graphical and modelling topology of networks.

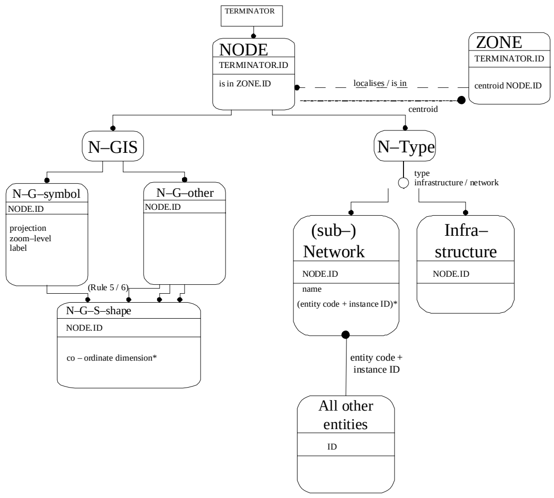


Fig. 31. GTF Example: Node

N–GIS (=TERMINATOR–NODE–GIS) is the container of graphical node information like projection, zoom–level of the co–ordinates associated with NODE in the entity N–G–S–shape (=TERMINATOR–NODE–GIS–SYMBOL–SHAPE). N–G–S–shape is a container of lists of co–ordinates describing the shape of a symbol used to display NODEs.

Nodes can be associated with ZONEs.

The function of a NODE as an element in the topology description of a supply side model makes a NODE of type “INFRASTRUCTURE”. If the NODE is used as an aggregation container of other entities, then the NODE is of type “(sub–)NETWORK”. And one uses this kind of NODE to “zoom in” and to “zoom out” of a NODE in order to see its internal structure.

The English definition (in a computer science context) is:

> zoom:
>
> <graphics> To show a smaller area of an image at a higher magnification ("zoom in") or a larger area at a lower
>
> magnification ("zoom out"), as though using a zoom lens on a camera.

In the context of this specification the definition above is enriched by the concept of “zooming in” to show further **topological** detail (not only graphical detail) associated with the NODE. But the basic idea of showing more detail or hiding it stays the same. The further detail a (sub–)NETWORK can associate with a NODE is, that the NODE is made up of other entities, e.g. a group of NODEs and LINKs that describes a railway station, an airport or generally terminals and their access and egress points as well as their ‘turns’ and ‘changes’ between LINKs (or LINKs of different modes!).

The entity ID's of all entities that a NETWORK instance comprises are migrated to that instance the NETWORK entity, such that the NETWORK instance becomes the grouping of all the entity instances that make up a specific (sub–)NETWORK.

Excourse "Information needed by Models"

The transport data that is covered is primarily that which is used in strategic transport models. Thus it covers interurban, regional or international travel on all transport modes for both passengers and freight. It does not cover detailed local traffic issues, such as the representation of road junction geometry.
In order to manage data, we must understand its basic characteristics. Data can be thought of as a symbolic representation of facts with meanings. A single meaning can be applied to many different facts. For example, the meaning "product name" could be applied to numerous combinations of alphabetic characters. A fact without a meaning is of no value. However a fact with the wrong meaning can be disastrous, as has been found by some international firms who realised that their standard product name had unfortunate connotations in the local language where they were marketing. Therefore, the focus of data management must be on the meaning associated with data.
Information can be defined as an aggregation of data for a specific purpose or within a specific context. Thus, the strategy to manage the information resource must focus on managing the meanings applied to facts, rather than attempting to control or limit the creation of information.
For further details we refer to the paper "Data needs for modelling" presented at the CONCERTO workshop 22/23 Oct. 1998, Brussels. In this paper the tasks, policy & modelling needs:

1 Policy needs and modelling
2. Queries and data needs for modelling
2.1. Generation
2.2. Distribution
2.3 Mode choice
2.4. Assignment
2.5. Interdependency of the modelling steps
2.6. Other models
2.7. Supply side
3. Resume

are discussed.

2.6.3.1.3 ETIS External Interface

Once ETIS is up and running, a further enhancement is already envisaged:

all GTF Translators can be linked together (and new translators can be written) to provide a homogenous I/O interface between the core of the system and the outside (external data & model providers etc.).

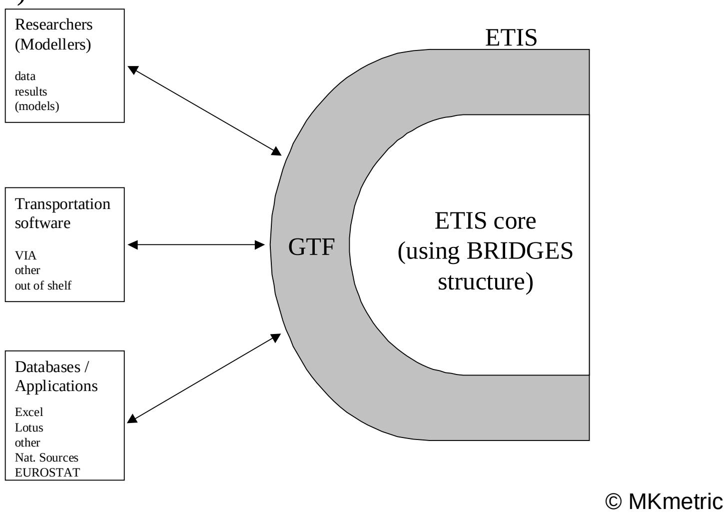


Fig. 32. GTF ETIS External Interface

To drive forward the automation process other supporting links have to be developed.

2.6.3.2 Supporting Automation

In order to be able to run, models generically a “language” for this purpose will be defined. This “language” will be incorporated into the GTF exchange format. It will allow to run models and retrieve the requested results automatically, without having to contact a model provider personally by defining a project to compute the results.

This will allow ETIS an effective use of external and internal models.

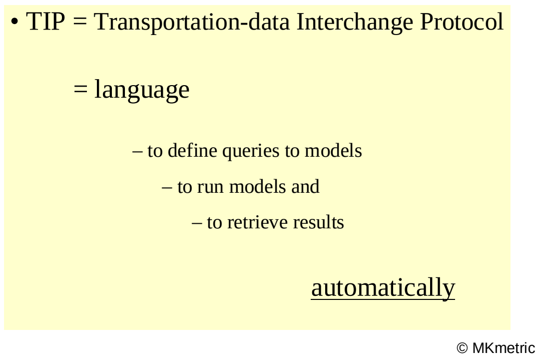


Fig. 33. Supporting Automation

Figures 10 - 17 established the specification for a "Generalised Transportation–data Format". Now a specification concerning the available commands to a user’s workspace is required. These commands will be attached to a GTF file and will enable a model provider to process the GTF data file so that the requested answers are computed. This is necessary, because a GTF data file alone doesn’t contain any information on what shall be done with the data. This is where TIP is necessary. TIP is a generalisation of "usual" commands (queries) to a transportation model. The development of TIP is based on the classic four step transportation model: generation, distribution, modal split and assignment. Within these four stages, a number of commands (independent of the actual model or the model's philosophy) can be issued to the model in order to produce intermediate data or final model results. These results can then be filtered through a filter defined in a TIP command file that is attached to a GTF data file. The filter extracts the data relevant to the user's query out of the model results and notifies the user's workspace that the requested results are available at the model provider.

2.6.3.2.1 Classification of possible queries

The categories of possible (transportation–)information exchange are:

pricing policies
regulatory policies
investment policies
co–operation of models

The following types (per category) are feasible – also ensuring that meaningful results could be produced:

modification of model input

input modification, e.g.
proportional modification of a variable's value on a whole network or a specific sub–set (for pricing policies, regulatory policies)
modifications of networks (for investment policies)
output queries, e.g.
modal split effects (e.g. high speed train vs. air; alternative i vs. alternative j)
generation and distribution effects on airport choice results

communication between models

Model provider A → MKmetric: output of model A (passenger movements or OD–flow matrix) as input to VIA
MKmetric → Model provider A: output of VIA (modal split matrix) VIA as input to model A

In respect to scenario definitions and future projections the described options fit into the following framework (please refer to "TRANSPORT RESEARCH APAS – Transport strategic modelling" for further information):

Please refer to Fig. 16.

For each of the components in the last level of the hierarchy two commands must be available:

Input modification:

1. explicit change of variable values, e.g. variable X = 100

2. functional change of variable values, e.g. variable X = (variable Y * 2) + variable Z (all mathematical standard operators and functions are allowed for manipulation, e.g. log(), sin(), +,–,*,/, exp() etc.)

Output query:

3. output matrix to be calculated, e.g. modal split for all available modes, assigned road network – Germany
4. definition of extracted variables, e.g. modal split of mode road and air, travel–time on link 1152, travel–time of shortest path between zone 51 and zone 894

The variables available in 1., 2. and 4. are the attributes of entity instances defined in the GTF data model, e.g. PHYSICAL_SPECIFICATION (of LINK 34254) – flow.

2.6.3.2.2 TIP commands

The commands needed are split into two categories:

manipulation of variables (selecting & setting / updating)
creating, requesting matrices (selecting & calculating)

These categories are based on the usual structure of transport models:

(See *Fig. 17.*)

The variables available for manipulation are those defined in the GTF data model, e.g. “entity FACTOR – INSTANCE 34923 – Population Class 79” for a *single* manipulation or “entity FACTOR – Population Class 79” for manipulation of *all* instances. The semantics for the manipulation commands is based on SQL, because the manipulation of GTF variables is a manipulation of relational data. **The manipulation commands** (i.e. manipulation of model input data) **always refer to data already located at the model provider.** (To manipulate data located in the GTF file doesn’t make sense, as the result of the manipulation can be computed beforehand, at the user’s site.) The commands have the following syntax:

```
UPDATE <entity>.<ALL|SINGLE> SET <variable>=<value>
UPDATE <entity>.<ALL|SINGLE> SET <variable>=<function>
UPDATE <matrix>
```

Where ‘function’ is a mathematical function of any variables in the GTF data already at the model provider.

A user can either modify single data elements (e.g. travel–time on link between node 42873 and node 42192 multiplied by 1.2) or lists of data elements (e.g. all the travel–times in a network multiplied by 1.1).

<matrix> is one of those specified in the following paragraph.

The requests for calculation are related to the usual phases that a transportation model comprises: generation (production / attraction), distribution, mode choice (modal split), traffic conversion, route choice and assignment (see *Fig. 17*). The request commands (i.e. model output data) are introduced by the keyword ”CREATE”
Command syntax: “CREATE <matrix> {MODE|PURPOSE|SEGMENT|PRODUCT}”
Followed by the these keywords:

KEYWORD	Computed / Resulting Matrix contents
GENERATION	Matrix: ZONE (x PURPOSE or SEGMENT / PRODUCT) Cell contents: ZONE number of TRIPS or amount of FREIGHT
PRODUCTION/ ACTTRACTION	Matrix: ZONE x ZONE (x PURPOSE or SEGMENT/ PRODUCT) Cell contents: number of TRIPS or amount of FREIGHT
DISTRIBUTION	Matrix: ZONE x ZONE (x PURPOSE or SEGMENT/ PRODUCT) Cell contents: number of TRIPS or amount of FREIGHT
MODAL SPLIT	Matrix: ZONE x ZONE x MODE (x PURPOSE or SEGMENT/ PRODUCT) Cell contents: amount of FLOW / PERCENTAGE (trips / tons)
TRAFFIC CONVERSION	Matrix: ZONE x ZONE x MODE (x PURPOSE or SEGMENT/ PRODUCT) Cell contents: number of VEHICLES
ASSIGNMENT	(loaded) network: NODE x NODE Cell contents: LINK attribute(s)

The keyword defines which output matrix shall be computed and transmitted (after filtering) back to the user. The specification of MODE, PURPOSE, SEGMENT / PRODUCT is optional. If one is specified it must follow the keywords preceding, e.g. ”CREATE DISTRIBUTION BUSINESS”.
The output filter is defined with

```
FILTER <matrix> <variable 1> ... <variable N>
```

The meaning of this line is: ”Filter from the output matrix <matrix> the variables <variable 1> through <variable N>”. Where <variable> is the fully–qualified name of an entity attribute.

2.6.3.3 Supporting Interface

This chapter will specify in a little more detail the elements needed to the ETIS user interface.

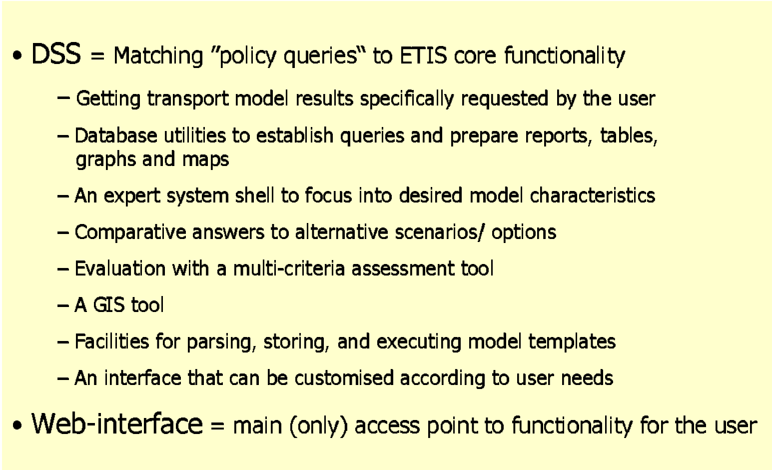


Fig. 34. Supporting Interface

The main elements are a Decision Support System and a web interface.

2.6.3.3.1 Decision Support System - DSS

A Decision-Support System (DSS) has been developed as part of the BRIDGES research project. It provides a set of tools that facilitate navigation of the European Transport Information System (ETIS) according to specific needs of a user.

DSS allows to select or formulate questions, transforms them into a series of calls to transport models and databases, and displays the results through its own routines (as reports, tables, graphs, maps) or through other ETIS modules.

The DSS can be used also as an independent module, since it has its own expert system shell, a multi-criteria assessment, database management capabilities, and GIS utilities.

The options available to DSS users include:

Getting transport model results specifically requested by the user
Database utilities to establish queries and prepare reports, tables, graphs and maps
An expert system shell to focus into desired model characteristics
Comparative answers to alternative scenarios/ options
Evaluation with a multi-criteria assessment tool
A GIS tool
Facilities for parsing, storing, and executing model templates
An interface that can be customised according to user needs

Modellers can make part of their models' functionality available to outside users by preparing DSS Templates. A template is an ASCII file that contains descriptions with keywords used by the DSS in order to parse, analyse and store the defined methods. For each specific user question the modeller prepares a method, which can be registered in the DSS database. Examples using MKmetric's VIA model results have been prepared to demonstrate the template concept.

The two main elements of the DSS are:

Activity: Decision-Support System

A key element of the ETIS Core Interface (CI) is its user friendliness. The DSS for making queries and interpreting results will enjoy similar user-friendly characteristics. It will be programmed as an independent application properly linked to ETIS Core Interface and other ETIS elements. The own CI programming language and functions will be tested (when ready) to facilitate the best possible integration of DSS into ETIS structure. DSS menus will allow the user to formulate specific questions and information requests to ETIS system, and transform them into a series of calls to transport models and/or databases.

The DSS will facilitate the following options:

(1) Initiation of transport models more appropriate to the functionality requested by the user;

(2) Database utilities to establish queries, updating, obtaining reports, etc;

(3) Establishment of expert rules to estimate missing data;

(4) Comparative answers to user questions relating to alternative scenarios/options; and

(5) Evaluation of model results according to multi-criteria models (e.g., should I recommend this road to be build this way?).

The DSS will be a combination of a database manager and an expert system. Options (1), (3) and (5) will be provided through an expert system module that will know, depending on the question, which model data to use, how to fill in data gaps, and what kinds of results to display to the user.

Activity - Expert System (ES)

The main task of the Expert System (ES) is to analyse user's queries, decompose them into sub-queries for passing to other modules, and combine results into a meaningful form for the user to understand. In doing so the ES will apply default values, educated guesses and qualitative criteria to fill in any missing data required for running other modules.

The ES will be part of the DSS and thus, will be accessible through ETIS Core Interface, or directly by its own user-friendly interface.

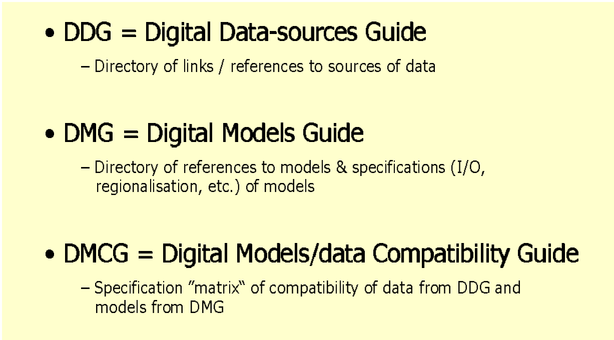
The ES will have the following layers: (1) Query Manager that interprets user queries, (2) an ETIS guided navigator that calls the appropriate modules and/or inspects the corresponding parts of its knowledge base, (3) a Default Manager that applies rules in order to find realistic assumptions for missing data, (4) an Evaluator that assembles results from the diverse sources above mentioned to be used in a multi-criteria model.

2.6.3.3.2 Web interface for the User Interface of ETIS

The Usability-gap of current user interfaces (UI) and an optimal UI requires re-thinking and re-designing of a central access point to the functionality that a user wants, therefor the following is needed (rather planned): Transparency, friendliness, interactivity; Exact definition of the UI (user interface) and definition of the best implementation strategy – local application or web-browser based with Applets contacting a (possibly distributed) server or the external resources (data, models) ?.

2.6.3.4 Supporting Guides

The ETIS supporting software system will need some guidance to be able to optimally aid a user in finding and using the necessary data and model to solve the user's queries.



© MKmetric

Fig. 35. Supporting Guides

2.6.3.4.1 Digital Data sources Guide - DDG

The Digital Data Guide (called DDG) was implemented in BRIDGES. Its general objective is to provide the European deciders with a comprehensive view of the transport data available in Europe, with a focus on digital data (databases and GIS). Special attention has been paid to the fact that this tool must be coherent with the other components of BRIDGES, particularly the «DSS».

The DDG is an interactive guide of the existing (and emerging) digital information products dealing with transport with European interest. This guide should be easily updateable. It will be multi-entries, describing in detail source references, formats and metadata (including confidentiality and cost). Particular attention will be paid to information networks involving several partners, such as World span.

It was co-ordinated with INFOSTAT, MESUDEMO, OD/ESTIM, COMMUTE and ASSEMBLING regarding the information requirements by various user's profiles and maintenance of the system.

The objective of the DDG is the assessment of present and future data sources availability, including accessibility criteria (format, cost, confidentiality, software, acquisition). The DDG does not contain data itself, but accurate source references. Under a user-friendly environment, the DDG will guide the user through a multi-entries tool describing: references/sources, formats and any other key relevant parameter

needed to evaluate the interest and the value of each European source of information. It will be linked to the Decision-Support System.
The DDG will be closely related to European user's needs and purposes. In this context, contacts with Eurostat, Gisco, DGVII and Infostat will be assured.

2.6.3.4.2 Digital Model Guide - DMG

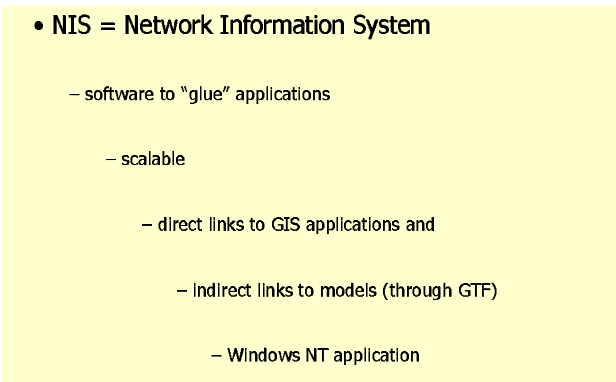
The DMG - the Digital Model Guide, which is analogous to the DDG but focuses specifically on models - in a much more detailed manner than in the DDG. The DMG would include detailed information about models, their requirements and their scope etc. With this the user could chose much more appropriate models for his query.

2.6.3.4.3 Digital Model / data Compatibility Guide - DMCG

The DMCG – Digital Model / data Compatibility Guide - which would state which data and models are compatible and to what degree, something which can't be done using only the DDG and the DMG. This would allow the system (actually the DSS) not only to guide the user to the data and the model, but to present a list of models & data fitting best to the query. A user could concentrate on finding data & models that answer his question precisely thus only using minimal data and model resources.

2.6.3.5 Supporting Software Communication

Ongoing research and development in the G.I.S sector is moving towards major openness, scalability and Internet compatibility, as well as towards major user friendliness for Desktop Mapping options. On the other hand, commercial transport models enjoying G.I.S. options are emerging (such as TransCad by Caliper) and GIS app. well linked to sophisticated DBS app. with utilities able to read any kind of data format (such as Geomedia). Database managers, through ODBC and further Internet JDBC extension, are in process of achieving data format exchange and, to same extend, query message communication through SQL protocols.



© MKmetric

Fig. 36. Supporting Software Communication

Network information System - NIS

Generally speaking, since software users become more software demanding, there is a natural move towards open systems and public formats. No single application can solve optimally all requirements, and well linked open platforms of CAD, DBS, GIS, DM and Transport Models are needed to carry on specialised works with enough productivity. Already existing closed and universal applications with proprietary data formats will tend to be substituted by a more or less large list of specialised intercommunicated applications (or even simple software components, stand-alone routines and controls) open to share data with other applications and to be driven externally, exchanging command messages.

In conclusion, the transport planning and management sector ideally requires a specific combination of CAD, GIS, DBM and DM Desktop Mapping open in order to support information systems in a proper manner. BRIDGES Core N.I.S. Utilities are defined in this frame.

CAD app. (such as AutoCAD or Microstation) facilitate the digitalisation of physical networks and reference environmental and territorial elements, with additional G.I.S. oriented routines (programmed of AutoLisp or UCM-DML, or in general in Visual Basic, or just assembled into G.I.S. sister applications such as Geographics-Microstation) simple graph databases can be supported.

G.I.S. more sophisticated applications (such as ArcInfo or MGE) have specialised network-modules with many graph-oriented routines (despite incomplete and in many cases sub-optimal). They provide with complex geographic routines for projection change and adjustment, zone intersection, super-imposition etc. as well as raster (such as satellite images, scanners) integrated processing.

DM applications (such as MapInfo, ArcView) enjoy basic G.I.S. options and provide easy graphic display of data and results. They tend to be open to external applications, such as Microsoft Excel or Access, where data and basic calculations can be more properly handled, or to complementary G.I.S. app.

DBM (such as Access, Dbase, Oracle...) are always needed to store and manage increasingly large data sets coming from many different sources. Neither CAD plus G.I.S. routines, nor G.I.S. or DM application are optimal for managing large databases.

Transport Models will tend to decentralise CAD, GIS, DBS and DM utilities to external app., focusing instead on highly sophisticated modelling routines and algorithms. Even some mathematical conventional processes in most cases will be handled by mathematical packages (such as Mathematica and others) or components. (I.e. links to GTF)

Conclusion:

BRIDGES research should be followed by ETIS research in these complementary directions:

One the one hand it defines an specific format for storing and transferring transport topologies (GTF), which integrates elements from CAD formats (DXF, DGN...), GIS formats (...), DBS (MDB, DBS...) and DM (SHP, MAP-TAB...) in a transport-oriented manner.

On the other hand, specific routines and utilities devoted to manage network-oriented databases (and importing/exporting to GTF) are being developed within BRIDGES Core. Similarly to GTF, these routines extend, in a transport-oriented direction, CAD, GIS, DBS and DM routines.

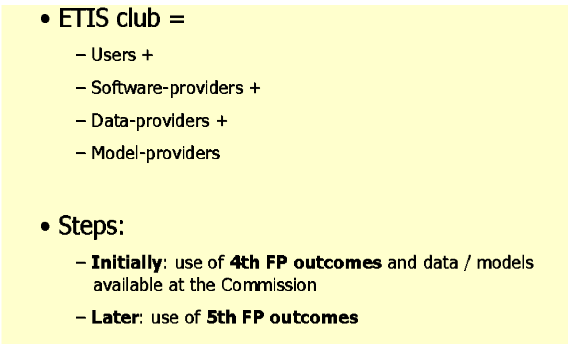
A user-interface able to manage transport graphs with high productivity (and integrate external applications) will be developed.

2.6.4 Initial Implementation & future of ETIS

In this chapter we want to outline a path to put ETIS into action. Starting from an initial implementation we point out the future potential and wide spread ramifications of ETIS. Finally, we shed some light on the technical network issues.

To start-up ETIS we suggest an initial implementation in a pre-selected environment. Here we can deal with and solve specific problems so that ETIS can become the best chance to grow in an open environment with qualified access. Like in the BRIDGES project we suggest EC-DGVII, EUROSTAT and EIB as initial users (and final customers) of ETIS, i.e. the ETIS club.

As soon as ETIS is working properly other EC-DGs and the ministries & public institutions of the EC member states should join the ETIS club and can be used as multiplicators. Again some time will be needed to endorse all new requirements and ideas, but as ETIS is conceptually flexible this process will happen fast. Finally, any user and provider can join the well established ETIS club and an independent institute should take over the task of operating ETIS.



© MKmetric

Fig. 37. Initial Implementation of ETIS

Below we summarise the development sequence of ETIS and the 'club' (= people using ETIS):

1. (Default) step: usage of

   consortium partners' modelling software

   free application executables

   models • GIS • EUROSTAT

and the further development steps of models, data, networks and software can proceed as follows:

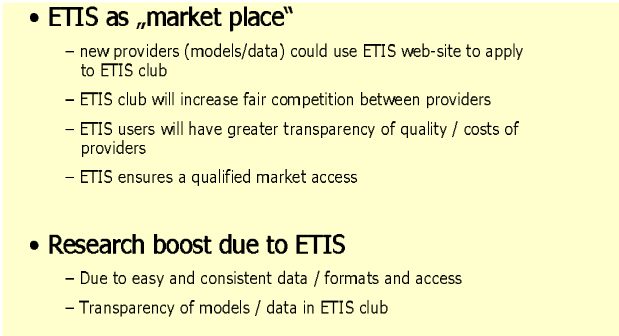
2. step: **tasks funded by the EC at 100%.**
3. step: **tasks funded by the EC at 50%.**
4. step: **as parts of future EC tasks and projects.**
5. step: **tasks contracted by the member states endowed by then (past, present, future projects, data etc.).**
6. step: **all qualified users and providers on a free basis from the outside market.**

To ensure the future of ETIS it shall not be a closed subsidised entity. Moreover, it is essential to ensure self-cultivating structures open to any kind of user, data, model, analysis and GIS tool provider. Of course one should take the chance from the beginning to establish qualified access rules, e.g. by a scientific committee, to ensure a high level of service, science, transparency and validated work. The access to ETIS should not be a matter of the entrance fee or user charges it should rather be the quality provided and the reputation & seriousness of users and providers.

Therefor ETIS should be established as a virtual market place where decision makers represent the demand side and service (data, model etc.) providers the supply side. It should be run by an independent institution, which is neither in conflict with politics nor has any commercial interests. If it appears that profit will be made then there should be the guarantee that the institute will finance EC-wide surveys or other data collection, harmonisation or homogenisation & ensuring compatibility tasks (to support the self-cultivating activities). In addition, one will have to ensure that the administrative tasks are as low as possible by freezing the budget and that a maximum of services the institute needs to run and maintain ETIS (- maintenance of the infrastructure elements, security, quality aspects and accounting procedures, electronic accessibility etc.) is contracted externally by market prices.

Standard maintenance topics will have to address the ongoing upkeep of the technical network aspects of ETIS like the servicing of all the servers and the permanent availability of the system, the regular checking that all members of the ETIS club continually fulfil the quality requirements and standards, keeping the list of members up to date, dealing with surprising events like fire, water outbreaks and all unforeseeable other events, the accounting of the usage of external provider (or internal services between the DGs) services and reimbursement of utilised resources and lastly constant implementation of the security levels needed by different users and providers.

Further maintenance topics will have to deal with updating and improving the guides (DDG, DMG and DMCG), enhancing and detailing GTF, adapting the exchange format of GTF from GESMES to other formats like XML and last but not least increasing the user friendliness of the user interface by implementing the access to all functionality of ETIS through a web interface, i.e. a web site (- for example by programming Java Applets that connect to the nearest ETIS server and run the necessary transactions). These are by no means complete lists, but they already show the level of detail to be dealt with on an operational basis once ETIS is up and running.



© MKmetric

Fig. 38. Self-Cultivating ETIS Structures

Implications / ramifications of GTF, BRIDGES and ETIS

The impact of the ETIS research and resulting bridging software will have many ensuing scientific & commercial and practical ramifications:

1. the compatibility between application specific files resulting from ETIS will allow users to stay in their known environments (workspaces) and use new software without wasting time having to learn the new software usage for importing / exporting. This will be done by bridging software and the GTF translators.
2. people dealing with problems appearing in different working areas can exchange information, e.g. analysing side effects when changing from a higher to a lower aggregation level
3. synergetic effects can result from the possibility of transferring knowledge between systems and points of view
4. the number of software, data, model licenses that will have to be purchased will decrease because of the centralised way the software and databases can be accessed using ETIS
5. a user can pre–select scenarios and identify those which really are relevant to his problem and then can be worked out in detail by human experts (consultants)
6. ETIS will give a boost to the availability of data in Europe; easy access to data will become as obvious in Europe as it already is in the United States of America
7. it will be possible to compare different models’ results (and their quality) as they can be used on the same data(–base)

8. model users won't always have to (re–)create their own databases over and over again like in the past, but will have access to standard data(–bases)
9. data–(bases) will gain in quality as time passes, because the data providers will have an incentive to update their databases regularly and properly, since only the "good" databases will be used and in addition good databases work out by master or ph.d. students can be made accessible
10. data providers, consultants and software developers will want to join the ETIS–'club, because it is a great way of making oneself known to the public (or the ETIS community) and therefore being a good selling platform (something like an open and free market)
11. countries which don't have a transport planning system can have access to the existing transport knowledge and use whatever suits their needs, i.e. they will have the possibility of planning transport infrastructures, something not done up to now
12. the fast pace at which telecommunication and telematics are developing in respect to the pace at which models are developing, suggests that within a few years users will be able to use even remote (call–in) models interactively
13. since the GTF specification belongs to the EC, it will have the possibility to ensure the technical quality, i.e. that the service providers linked to ETIS meet some standard hardware or availability requirements etc., or model quality, i.e. that the model providers ensure some standard of quality of their models and results resulting in a high performance (= speed, requested detail, ease–of–use, availability of an e–mail complaints box etc.), because only the requested results are transferred in the appropriate detail. The quality of the data could be ensured by EUROSTAT or national public services like national statistical offices.
14. a standard for transport projects can be defined by databases, models and comparative result representation for any user–group, so the results of different projects can be compared
15. researchers from different countries can work on the same database and exchange knowledge about their findings
16. some ETIS' links might become obsolete, as application developers open their system to import and export their application's data or directly use the GTF specification, but most ETIS components, e.g. the 'Decision Support System' (DSS), GTF etc., won't become obsolete, because the application developers just will provide simple import / export routines, without the power and general applicability of ETIS' components

17. once GTF and ETIS are in place and in use, user's needs will make GTF grow to accommodate the new requirements
18. specific confidential workspaces can, will and have to be defined to also ensure different security levels, once ETIS gets opened up
19. users will request new models or combination of models, which previously could have been denied by the consultants, because of lack of transparency on the supply–side of the business
20. transport companies will be able to find new market niches to supply with their service(s)
21. to ensure a high level of technical and model quality the institution running ETIS could open an e–mail complaints box for the users, reporting difficulties accessing or using some service provider etc. or other problems, which could be monitored, resulting in formal complaints to the service providers to raise their standard of quality
22. with the control over ETIS by an independent institution one can ensure that new service providers will first have to be able to cope with test–suits, before allowing their complete entrance to ETIS, ensuring a high quality level of service (enough help support for their services, validation of their data / results and models etc.) through ETIS
23. with the possibilities of automation using TIP a very important (and main) aspect of the user-friendliness of the system can be developed, as with the automation using TIP, the system can relieve the user from many tedious and unnecessary (from the user's point of view) tasks.
24. users and providers subscribing to / entering the ETIS club will have to pass quality (qualifier) procedures. These can later also be used for maintenance duties like the updates of guides, queries, GTF / TIP etc. The same holds for users and providers which are already in the club as they will be interested in updating their services and to include new queries or demands into the DSS
25. all these effects will have a vigorous impact on research in this and other fields

These are just some interdependencies one can think of. Of course there will be many more.

Fig. 39 depicts the vision of ETIS as a system that allows users from anywhere (at least the whole of Europe) to access ETIS and to make productive work over the Internet.

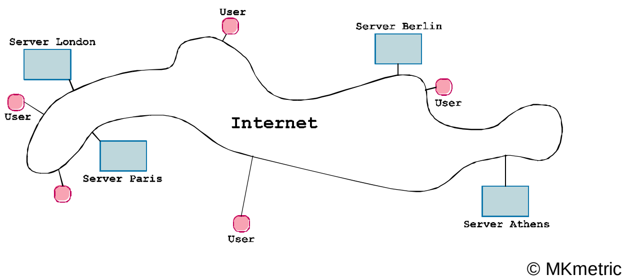
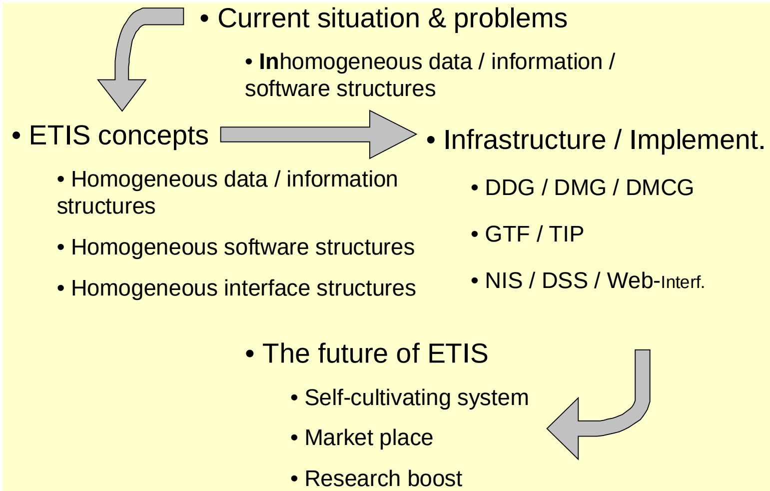


Fig. 39. ETIS club Intra- / Internet Structure

It would exceed the focus of this paper to start the discussion of what kind of technical network should be used, how the servers have to be structured or whether it makes sense to follow a centralised or de-centralised strategy. But as the points raised show, we already thought of this so that we just want to state at this point that we recommend a high-speed backbone network with decentralised server architecture and a central backup, security and accounting system.

2.6.5 CONCLUSION / SUMMARY

The goal of ETIS is to support policy-makers during the whole process of “Policy Scenario” formulation through to making the actual decision. To be able to do this , the decision maker nowadays has many obstacles to overcome because the data, the data structures, the information contained in the data and last but not least the software and software structures are mostly (mainly) in-homogeneous and the formats proprietary. This makes the task of finding sound (and comparable) information to base a decision on and the comparison of such information (e.g. scenario results from models) nearly impossible due to the virtual impossibility of exchanging data / information from one model / software to another and the already mentioned in-homogeneity of all the (necessary) intertwined structures.



© MKmetric

Fig. 40. Summary

In order to develop a system to support a decision maker in a user-friendly way the concept of ETIS were developed which will guarantee homogeneous data, data structures, information contained in the data and software and software structures. The main concept is a number of guides which serve as directories of available data, models and compatibility between both and between the models, software to glue applications together that weren't initially developed to communicate with each other, a homogeneous data (information) exchange format and a user-friendly interface. Together, these components "are" ETIS.

The ETIS structures were conceived to be "self-cultivating", i.e. only minimal effort from outside (e.g. the European Community) will be needed to maintain ETIS and for ETIS to prosper and grow. The "ETIS club" will make sure of this without having to explicitly make the effort, because of the "invisible hand" of fair regulations, quality controls and market competition.

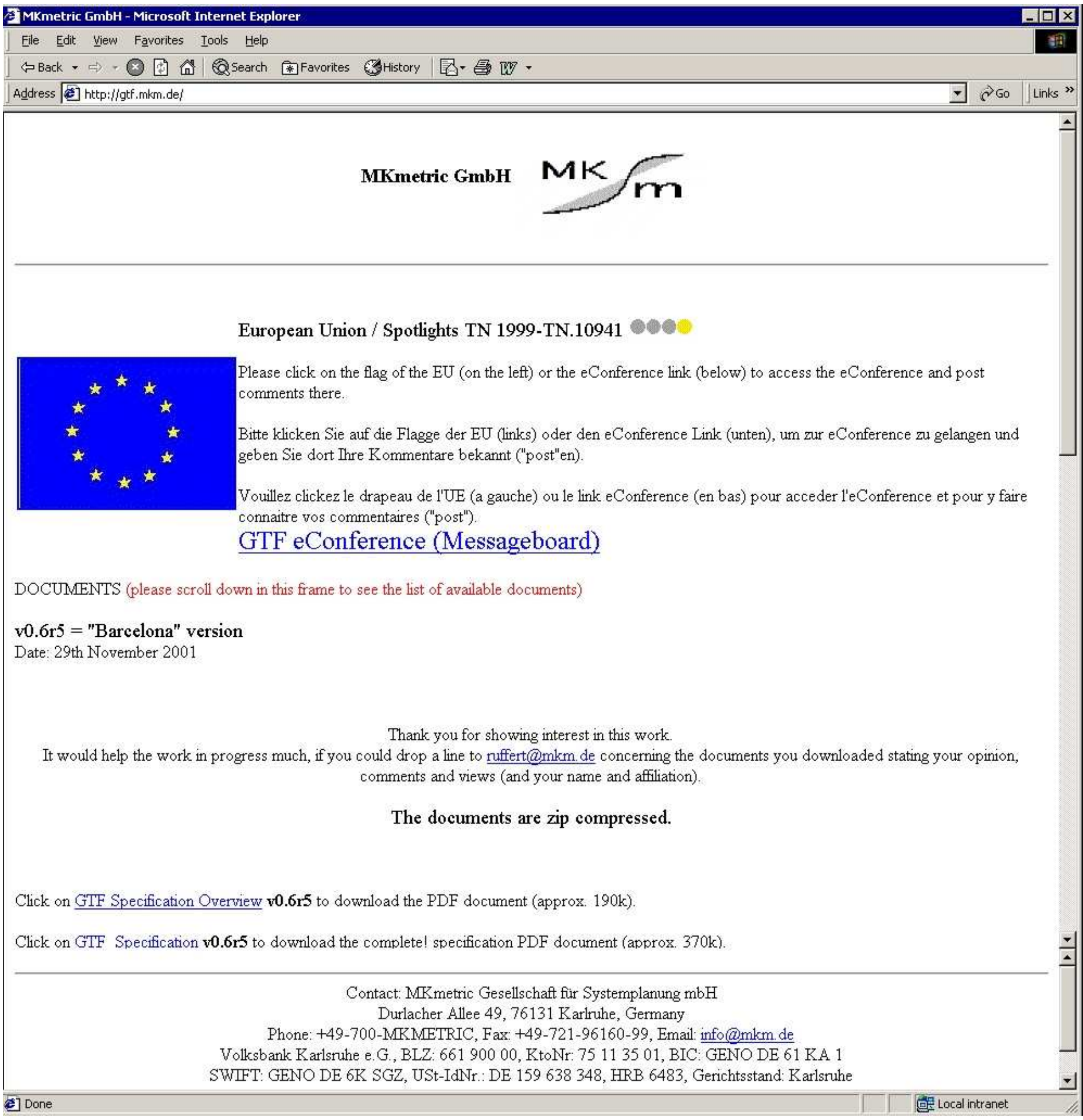
2.7 SPOTLIGHTSTN – FINAL CONFERENCE

See the conference proceedings provided by the co-ordinator.

2.8 SPOTLIGHTSTN – WEB SITE

MKmetric installed a "GTF Web Site" http://gtf.mkm.de.

The following is a screen shot of the site's main page.



This site provides all the latest / current versions of the documentation concerning the GTF Specification as well as other activities documents. Additionally it is the portal to the GTF Web Forum which can be reached by clicking on the flag of the EU.

The dissemination of the main documents concerning GTF (GTF Specification Overview, GTF Specification, GTF_adhoxXML and GTF-TIP) was logged and the statistics are as follows:

```
Document = Number of downloads:

GTF Specification =  39
```

GTF Specification Overview = 47

GTF ad hoc XML = 23

GTF-TIP = 24

Downloaded from 71 different servers, namely

129.217.191.59
160.40.60.54
193.130.171.33
193.170.124.138
194.134.164.188
194.154.214.69
194.200.93.100
194.25.187.65
195.112.131.56
195.251.234.72
203.141.89.174
205.205.212.26
212.209.219.223
212.92.36.137
61.150.43.20
62.152.67.38
62.252.64.5
66-44-60-163.s163.tnt4.lnhva.md.dialup.rcn.com
66-44-92-84.s592.tnt6.lnhva.md.dialup.rcn.com
66.92.57.131
a1as10-p142.mch.tli.de
adsl-130-76.wanadoo.be
ash.meap.co.uk
bw2-114pub234.bluewin.ch
c2.sll.se
cache-hki-2.inet.fi
cache-hki-4.inet.fi
conker.meap.co.uk
crawl1.googlebot.com

crawl2.googlebot.com
crawl4.googlebot.com
crawl5.googlebot.com
crawl6.googlebot.com
crawl7.googlebot.com
cx44640-b.rsmt1.occa.home.com
DGEO157.GEO.SBG.AC.AT
dialup-64.157.53.121.Dial1.Washington1.Level3.net
diomedes.noc.ntua.gr
dsl092-057-131.lax1.dsl.speakeasy.net
dummy2.minvenw.nl
dwp.esri.com
dyn230-ras18.screaming.net
dyn237-ras15.screaming.net
F204.ifw-kiel.de
ffm2-t6-2.mcbone.net
fire.engineering.Virginia.EDU
fw.bane.dk
fw.kti.hu
gw3.telekom.de
gw-cro01.pgb.philips.com
hme0.icc.es
ilgpcr.ethz.ch
inktomi2-cam.server.ntl.com
inktomi3-ltn.server.ntl.com
ip658.boanxx2.adsl.tele.dk
JB.nsd.uib.no
mail.cmc.com.vn
mail.sdgworld.net
merkur.netpioneer.de
nas-user137.ictnet.es
nat-ph3-ext1.rz.uni-karlsruhe.de
net141-051.mclink.it
oan.ctt.dtu.dk
pc1.profitmediat.pro.fi
ru-lt12.inf.ethz.ch
sim6.inf.ethz.ch
THREE-CAMBRIDGE-CENTER-TWO-SEVENTY-FOUR.MIT.EDU
training6.caliper.com

```
user.itc.nl
user-uiver34.dsl.mindspring.com
vkk037.citg.tudelft.nl
```

Date: 5th February 2002

Conclusions

As can be seen by the number of downloads from the GTF Web Site response rate it is more successful than the response rate of the eConference and the web forum. This seems to be because experts and consultants etc. whose participation is crucial in the discussions concerning GTF are only willing to participate if their time they spend reading, commenting at workshops etc. is reimbursed which was not the case for the participation at the eConference and also not at for the web forum. The motto "time is money" is of course especially important for those working in the private sector.

The manner in which the downloads were easily retrievable from the GTF web site and the statistics above show that this form fits best for the transportation community.

2.9 ATOM: SCENES – GTF DEMONSTRATION

The spotlightsTN/GTF task co-operated with the ATOM project in its evaluation work concerning GTF. The questions for posed by ATOM were in the line of "What will it take to actually implement a GTF Translator and make a system GTF enabled?", "Are there any conceptual difficulties with the GTF Specification?", "How far can the current conceptual GTF Specification be used in defining a concrete example?" and "Can a demonstration be made of an application of GTF to some exemplary files from SCENES?".

These questions were useful during the last stages of rounding up the GTF work. For more information refer to the "ATOM REPORT".

2.9.1 EXAMPLE ENCODING OF SCENES DATA IN GTF-XML

Transport Supply – Network Infrastructure

```
<?xml version="1.0"?>
<!Encoded from MEPLAN SCENES FILE: new_scenes/base_final/utn.dat>
<!Translation into GTF-adhoc-XML format v0.6.r5>
<!Authors Paula Cuthbertson & Neil Raha>
<!Creation Date 03/12/2001>
```

```
<!Last Updated  07/12/2001 by PJC>

<!******************************************************************!>

<GTFDB>
  <!Definition of SCENES network:>

    <!Node Definitions>
    <!Terminator Nodes>
      <N id=”n_1” name=”node_1.0000” type=”1” >
        <TE id=”te_1” centroid_of=”zo_1” >
      </N>
      <N id=”n_2” name=”node_2.0000” type=”1”>
        <TE id=”te_2” centroid_of”zn_2” >
      </N>
    <!End of Terminators>

    <!Junction Nodes>
      <N id=”n_3” name=”node_2.0002” type=”3” >
        <JU id=”ju_1” >
      </N>
      <N id=”n_4” name=”node_92.32” type=”4” >
        <JU id=”ju_2” >
      </N>
      <N id=”n_5” name=”node_194.31” type=”4” >
        <JU id=”ju_3” >
      </N>
      <N id=”n_6” name=”node_92.0110” type=”13” >
        <JU id=”ju_4” >
      </N>
      <N id=”n_7” name=”node_194.0064” type=”13” >
        <JU id=”ju_5” >
      </N>
      <N id=”n_8” name=”node_4.0007” type=”11” >
        <JU id=”ju_6” >
      </N>
      <N id=”n_9” name=”node_129.0001” type=”11” >
```

```
        <JU id=”ju_7” >
      </N>
      <N id=”n_10” name=”node_15.3000” type=”4” >
        <JU id=”ju_8” >
      </N>
      <N id=”n_11” name=”node_176.3000” type=”4”>
        <JU id=”ju_9” >
      </N>
      <N id=”n_12” name=”node_15.0015” type=”13” >
        <JU id=”ju_10” >
      </N>
      <N id=”n_13”  name=”node_15.0002” type=”12” >
        <JU id=”ju_11” >
      </N>
      <N id=”n_14” name=”node_15.0003” type=”12” >
        <JU id=”ju_12” >
      </N>
      <N id=”n_15” name=”node_1.0078” type=”3”>
        <JU id=”ju_13” >
      </N>
    <!End of Junctions>
  <!End of Node Definitions>


  <!Link Definitions>
    <!Road Links - Car connectors>
      <L id=”l_1” comment_list=”cm_1, cm_3” starts_in=”n_1”
    ends_in=”n_2” direction=”0” type=”101” >
        <CO id=”co_1” avg_speed=”60” avg_capacity=”99999999” >
      </L>

      <L id=”l_2” comment_list=”cm_1, cm_4” starts_in=”n_3”
    ends_in=”n_4” direction=”0” type=”101” >
        <CO id=”co_2” avg_speed=”60” avg_capacity=”99999999” >
      </L>

  <!Channel Tunnel Links:>
```

```
   <L id=”l_3” comment_list=”cm_5” starts_in=”n_5” ends_in=”n_6”
direction=”0” type=”703” >
     <SEG id=”seg_1” >
   </L>
   <L id=”l_4” comment_list=”cm_5” starts_in=”n_7” ends_in=”n_5”
direction=”0” type=”704” >
     <SEG id=”seg_2” >
   </L>
   <L id=”l_5” comment_list=”cm_5” starts_in=”n_8” ends_in=”n_6”
direction=”0” type=”704” >
     <SEG id=”seg_3” >
   </L>

<!Truck Border links:>
   <L id=”l_6” starts_in=”n_9” ends_in=”n_10” direction=”0”
   type=”104” >
     <SEG id=”seg_4” >
   </L>
<!Road Ferry links:>
   <L id=”l_7” starts_in=”n_11” ends_in=”n_12” direction=”0”
type=”703” >
     <SEG id=”seg_5” >
   </L>
<!Road to Ferry links:>
   <L id=”l_8” starts_in=”n_13” ends_in=”n_11” direction=”0”
type=”704” >
     <SEG id=”seg_6” >
   </L>
<!Road Links>
   <L id=”l_9” starts_in=”n_14” ends_in=”n_13” direction=”0”
type=”1202” >
     <SEG id=”seg_7” >
   </L>
   <L id=”l_10” starts_in=”n_15” ends_in=”n_13” direction=”0”
type=”1202” >
     <SEG id=”seg_8” >
   </L>
```

```
<!End of Link Definitions>

<!Mode Definitions>
  <MO id="mo_1" name="mode_20" comments_list="cm_8"
specifies_mode="l_6,l_9,l_10" allows="v_,v_," type="3" >
    <RO id="ro_1" name="road_ShortTruck" >
  </MO>
  <MO id="mo_2" name="mode_21" comments_list="cm_9"
  specifies_mode="l_6,l_9,l_10" allows="v_,v_" type="4" >
    <RO id="ro_2" name="road_ShTruckFee" >
  </MO>
  <MO id="mo_3" name="mode_22" comments_list="cm_10"
specifies_mode="l_6,l_9,l_10" allows="v_,v_," type="3" >
    <RO id="ro_3" name="road_TruckDrive" >
  </MO>
  <MO id="mo_4" name="mode_23" comments_list="cm_11"
specifies_mode="l_4,l_5,l_8" allows="v_,v_," type="3" >
    <WA id="wa_1" name="water_TruckAccom" >
  </MO>
  <MO id="mo_5" name="mode_24" comments_list="cm_13"
specifies_mode="l_4,l_5,l_8" allows="v_,v_," type="3" >
    <WA id="wa_2" name="water_TruckUnAcc" >
  </MO>
  <MO id="mo_6" name="mode_27" comments_list="cm_12"
  specifies_mode="l_4,l_5,l_8" allows="v_,v_," type="3" >
    <WA id="wa_3" name="water_ShTrAccom"  >
  </MO>
  <MO id="mo_7" name="mode_51" comments_list="cm_14"
  specifies_mode="l_1,l_2,l_3,l_7,l_9,l_10" allows="v_,v_," type="3" >
    <RO id="ro_4" name="road_CarRide" >
  </MO>
  <MO id="mo_8" name="mode_52" comments_list="cm_15"
  specifies_mode="l_1,l_2,l_3,l_7,l_9,l_10" allows="v_,v_," type="3" >
    <RO id="ro_5" name="road_CarFeeder" >
  </MO>
```

```
  <MO id=”mo_9” name=”mode_53” comments_list=”cm_16”
  specifies_mode=”l_1,l_2,l_3,l_7,l_9,l_10” allows=”v_,v_,” type=”3” >
    <RO id=”ro_6” name=”road_CoachRide”  >
  </MO>
  <MO id=”mo_10” name=”mode_54” comments_list=”cm_17”
  specifies_mode=”l_1,l_2,l_3,l_7,l_9,l_10” allows=”v_,v_,” type=”3” >
    <RO id=”ro_7” name=”road_Car_busi” >
  </MO>
  <MO id=”mo_11” name=”mode_55” comments_list=”cm_15”
  specifies_mode=”l_1,l_2,l_3,l_7,l_9,l_10” allows=”v_,v_,” type=”3” >
    <RO id=”ro_8” name=”road_CoachFeed” >
  </MO>
  <MO id=”mo_12” name=”mode_61” comments_list=”cm_15”
specifies_mode=”l_3,l_7” allows=”v_,v_,” type=”3” >
    <RA id=”ra_1” name=”rail_TrainRide >
  </MO>
  <MO id=”mo_13” name=”mode_62” comments_list=”cm_15”
specifies_mode=”l_3,l_7” allows=”v_,v_,” type=”3” >
    <RA id=”ra_2” name=”rail_Train feeder” >
  </MO>
<!End of Modes>

<!Metadata for network Definitions>
<M>
  <!Link Attributes:>
    <LA id=”la_1” length=”20” associations=”l_1” >
    </LA>

    <LA id=”la_2” length=”35” associations=”l_2” >
    </LA>

    <LA id=”la_3” length=”48.805” cost=”110.33” time=”0.6”
    associations=”l_3” >
    </LA>

    <LA id=”la_4” length=”0.1” time=”0.5” associations=”l_4, l_5” >
    </LA>
```

```
    <LA id=”la_5” length=”0.01” time=”0.15” associations=”l_6” >
    </LA>

    <LA id=”la_6” length=”355.681” cost=”137.73” time=”13.75”
  associations=”l_7” >
    </LA>

    <LA id=”la_7” length=”0.1” time=”1.0” associations=”l_8” >
    </LA>

    <LA id=”la_8” length=”1.016” time=”0.0122” associations=”l_9” >
    </LA>

    <LA id=”la_9” length=”2.323” time=”0.028” associations=”l_10” >
    </LA>

<!End of Link Attributes>

<!Dimensions - ie measurement units>
  <DI id=”di_1” SI_unit=”1” prefix=”7” property=”length”
associations=”la_1, la_2, la_3, la_4, la_5, la_6, la_7” >
  </DI>
<!End of Dimensions>

</M>
<!End of Metadata>


<!Additional Comments>
  <COMT id=”cm_1” >
    2 way link
  </COMT>
  <COMT id=”cm_2” >
    1 way link
  </COMT>
  <COMT id=”cm_3” >
```

```
  zone = Bergenland, centroid = Eisenstadt
</COMT>
<COMT id=”cm_4” >
  zone = Niederösterreich, centroid = St. Pölten
</COMT>
<COMT id=”cm_5” >
  channel tunnel link
</COMT>
<COMT id=”cm_6” >
  cost units = ECUs
</COMT>
<COMT id=”cm_7” >
  time units = hours
</COMT>
<COMT id=”cm_8” >
  mode_20 = Light Goods Vehicle truck drive
</COMT>
<COMT id=”cm_9” >
  mode_21 = Light Goods Vehicle feeder
</COMT>
<COMT id=”cm_10” >
  mode_22 = Heavy Goods Vehicle truck drive
</COMT>
<COMT id=”cm_11” >
  mode_23 = Accompanied Heavy Goods Vehicle on ferry
</COMT>
<COMT id=”cm_12” >
  mode_27 = Accompanied  Light Goods Vehicle on ferry
</COMT>
<COMT id=”cm_13” >
  mode_24 = UnAccompanied Heavy Goods Vehicle on ferry
</COMT>
<COMT id=”cm_14” >
  mode_51 = Car Ride
</COMT>
<COMT id=”cm_15” >
  mode_52 = Car Feeder
```

```
  </COMT>
  <COMT id="cm_16" >
    mode_53 = Coach ride
  </COMT>
  <COMT id="cm_17" >
    mode_54 = Car for business
  </COMT>
  <COMT id="cm_18" >
    mode_55 = Coach Feeder
  </COMT>
  <COMT id="cm_19" >
    mode_61 = Train ride
  </COMT>
  <COMT id="cm_20" >
    mode_62 = Train feeder
  </COMT>
<!End of Comments>
<!End of SCENES Network Definition>
</GTFDB>
```