

5 May 2008



**Deloitte.**

**THALES**



# National Transport Information Incubator (NaTII)

## Enabling the power of information

Professor Peter Landshoff<sup>1</sup> and Michael Simmons, University of Cambridge  
Dr Rashmi Joshi, Deloitte and Touche LLP  
Charles Gill, Lockheed Martin UK Ltd  
Colin Waugh, Thales Research and Technology (UK) Ltd  
John Patman, Cambridge University and Thales

*This document follows from a 7-month pilot study commissioned by the Department for Transport and is copyright to the authors*

---

<sup>1</sup> Contact person; email [pvl@damtp.cam.ac.uk](mailto:pvl@damtp.cam.ac.uk)

This Report is the final deliverable to **DfT PPRO 4/54/9 INVITATION TO TENDER: INFORMATION INCUBATOR - PILOT PROJECT**, dated 17 August 2007. It is submitted by Cambridge University on behalf of the NaTII Consortium.

Although this report was commissioned by the Department for Transport (DfT), the findings And recommendations are those of the authors and do not necessarily represent the views of the DfT. While the DfT has made every effort to ensure the information in this document is accurate, DfT does not guarantee the accuracy, completeness or usefulness of that information; and it cannot accept liability for any loss or damages of any kind resulting from reliance on the information or guidance this document contains.

## Contents

1 Executive summary	9 Project identification
2 Background	10 Examples of projects
3 What will NaTII do?	11 Barriers
4 Data	12 Organisation
5 Who is NaTII for?	13 Costs
6 Evaluation	14 Implementation
7 Expertise	15 Acknowledgments
8 "My Journey"	

Appendix A: Research

Appendix B: The authors of this report



**Expertise** NaTII will offer help with

Data: security, quality, brokering access, synthesising, assessing quality

Ideas: through workshops, brainstorming

Tools: mashups, synthetic environments, open source availability

Experience: with technology, end-user interfaces, the environment, disabled needs, portfolio of past projects

Collaboration: team building, management, legal agreements, intellectual property

Access: to experts, industry, government, public agencies

Assessment: evaluation, costs, benefits

**“My Journey”** My Journey was the application the pilot study used to help design the NaTII process and demonstrate its viability. It was to help arriving passengers at Stansted airport reach their destinations by delivering real-time information through mobile phones. The project client was BAA. Among the benefits BAA said it received were

- An outcome in effectively a 4/6-week period – a saving of some 12 months over its normal business process
- Verification of a complex process to agreed dates
- A high level of confidence in achieving the development budget
- Effective reduction of front-end cost, and a range of expertise and knowledge and consultants that you just would not normally get.

**Project identification** Some projects will be driven by a perceived need for a new service, others by the existence of some capability. An important function of the NaTII team will be to seek out and talk with owners of data, to encourage people do things they did not know they wanted to do. Although one aim of NaTII is to reduce risk, it cannot be completely eliminated and those who propose projects must be willing to take calculated risks.

Because NaTII will achieve a rapid initial assessment, projects that are judged to have relatively little value will not be taken very far in the process. As competition for access to the NaTII process builds up, the Director will need advice on project selection from an expert advisory group. He or she will report selected projects to a programme management board, which will include representation from the funding body.

**Examples of projects** Several projects that might be subjected to the NaTII process were identified during the pilot study. They include reducing the environmental impact of travel to, from and at work; travel in connection with major events such as the Olympics, handling in particular crises in the transport system; improving the design of major new urban regeneration developments; and connected cars and roads.

**Barriers** The *Data Grand Challenge* identified several barriers to organizations collaborating to make use of their data. These related to locating sources of data, access, limitations on use, confidentiality and security. Others emerged during the pilot study. In particular, with each member organisation in a collaboration bringing in a set of lawyers and, with each lawyer feeling the need to contribute to the negotiation, it takes much time and patience to put agreements in place. NaTII will help with alternative contractual arrangements. It will establish different data access protocols and new and accelerated evaluation processes to verify value.

**Organisation** A collaborative unit should be formed, with each organisation seconding key players into NaTII. Each organisation through its parent company should enter into joint and several arrangements with NaTII in order to supply it resources and services. As the organisations will have access to their own integration facilities, such as computing, there need not be a dedicated facility. The collaborative unit selects a director, who is responsible to the principal funder. The Director should be perceived to be neutral and should carry some prestige. The Director should be supported by a core of staff covering the expertise listed above. The expertise of the staff should be broad rather than necessarily deep, and so one person will cover several of the items. A programme management board should include a representative of the principal funder.

**Costs and implementation** At least initially, NaTII will need to be funded by government; this will help

it to be perceived to be a neutral territory. Nearly all of the cost of NaTII will be people, though some further tool development will be needed.. Not all the projects put through NaTII will be judged to be viable and, because of the iterative nature of the evaluation process, those that do not get through to the implementation stage will take varying times to the decision that they be parked. On the basis of 24 successful projects over the initial three years of NaTII, the estimated budget is £7 million. As data and knowledge integration are central to government policy, the aim should be for NaTII to begin work on its first project no later than January 2009.

## 2 Background

Advances in information technology are removing technical and economic obstacles to the combination of federated data belonging to different organisations. This brings with it the potential to create innovative information services and applications of considerable public and commercial value. Yet there remain various barriers that impede development of such services, including the difficulty of determining the costs and benefits of data applications without an opportunity to experiment.

In 2006 the Department for Transport was tasked with leading scoping work for the Science and Innovation Ministerial Committee's *Data Grand Challenge*. The conclusions are that the potential benefits of better data use include

- More informed policies - better access to data improves the development, delivery and monitoring of policies.
- Improved public services - data sharing between central and local government, and their partners, to produce more efficient, personalised, customer-focussed public services.
- Better regulation - reduce duplicated requests to citizens and business by more efficient use and sharing of data already held.
- Foster innovation - through greater access public sector data.
- Protect vulnerable groups - better data sharing can better target services to those most in need and help coordinate work between agencies.
- Empowering citizens - by improving access to public sector information will help hold government more accountable.

As a result of this work, the Department produced a "thinkpiece" examining the case for a "data mashing" laboratory or forum that would provide an experimental environment for the development of innovative information services. The Government response to the *Power of Information* review committed DfT and the Chief Scientific Adviser's Committee to complete scoping work for such a facility.

On 17 August 2007, the Secretary of State for Transport awarded a contract for "Information Incubator - Pilot Project" to the Consortium led by Cambridge University. The contract was for a period of 7 months and was awarded on a fixed price basis. Work began on 5 September 2007. This report summarises the work. The key deliverable from this study was a specification for a project demonstrating the benefits, business and technical models of an "information incubator" to facilitate the development of innovative transport information services, exploiting data sources held by private and public sector organisations.

Our pilot study was conceived at the time that work began on the *Power of Information* report and it takes forward the recommendations related to exploiting new web technologies and data mashing. The initial concept was to study a facility that would **determine the value of information through experimentation**. The aim of the pilot project was to

- develop the business case for a facility;
- create the initial NaTII environment, including process and tools;
- put a transport-related application through the NaTII environment so as to test that the process works, and learn what form the final NaTII facility needs to take;
- build expert end-user groups.

The application on which we focused in the pilot study was "My Journey": a project in collaboration with

BAA to help arriving passengers at Stansted airport reach their destinations by delivering real-time information through mobile phones.

The pilot project was undertaken by a consortium of organizations whose culture is very different: the UK arm of Thales, a French company; the UK arm of Lockheed Martin, an American company; a multinational company, Deloitte; and a leading university, Cambridge. Our combined experience in large-scale systems integration has offered proven methods for efficient and effective collaboration. We have brought together expertise in

- data mashing
- web technology
- security
- networking
- data quality
- business cases
- innovation

Our concept of NaTII has evolved over the 7 months of the pilot project. It should be much more than a data-mashing laboratory. As we see it, NaTII will be a pioneering mechanism for innovation. It will encourage people do things they did not know they wanted to do, or could do, with data and new technologies, by rapid feasibility assessment. It will enable high-speed de-risking, reduce the time to market, and increase the certainty of delivery.

It is key that NaTII should carry prestige and neutrality, so as to give owners of data the confidence to engage with it.

Based on our knowledge of other facilities, and from discussion with more than two dozen major organisations over the course of our pilot study, we believe that NaTII will be unique. It will encourage public and commercial bodies, and academia, to collaborate and give them help with all the many issues they must address, by bringing in a wide range of attitudes and skills so as to enable the whole picture to be exposed. We have learnt how even large companies, with access to a very diverse range of expertise within their own organisations, derive enormous value through collaborating with others whose culture and general approach are different.

This report concentrates on uses of information in the general area of transport, but almost everything in it would apply equally to every other sector of the economy. Indeed, one very important benefit of the NaTII approach is to transfer expertise and best practice between sectors. Services based on exploiting data potentially have enormous world-wide markets, which the UK is in a strong position to exploit, because of its investment in escience. But only if it continues its forward-thinking approach.

### 3 What will NaTII do?

NaTII will be a pioneering mechanism for innovation. It will

- reduce risk
- cut cost of evaluation
- reduce time to market
- encourage better data handling
- encourage gathering more and better data
- inform and test government policy
- cross-fertilise expertise.

It will provide

- a neutral environment for collaboration and experiment
- a technical architecture to help create rapid mashups
- an expert network, bringing the right people together
- an evaluation process

A project that is subjected to the NaTII process is first subjected to an **evaluation** stage. In terms of the Technology Readiness Levels introduced by NASA and described in

[http://en.wikipedia.org/wiki/Technology\\_Readiness\\_Level](http://en.wikipedia.org/wiki/Technology_Readiness_Level),

this stage covers:

- Basic Technology Research
  - TRL 1 Basic principles observed and reported
  - TRL 2 Technology concept and/or application formulated
- Technology Development
  - TRL 3 Analytical and experimental critical function and/or characteristic proof-of concept
  - TRL 4 Component and/or breadboard validation in laboratory environment

If the evaluation stage concludes that it is justified, the project is then taken forward to an **implementation** stage, which covers the remaining Technology Readiness Levels:

- Technology Demonstration
  - TRL 5 Component and/or breadboard validation in relevant environment
  - TRL 6 System/subsystem model or prototype demonstration in a relevant environment
  - TRL 7 System prototype demonstration in a space environment
- System Test and operations
  - TRL 8 Actual system completed and “flight qualified” through test and demonstration
  - TRL 9 Actual system “flight proven” through successful mission operations



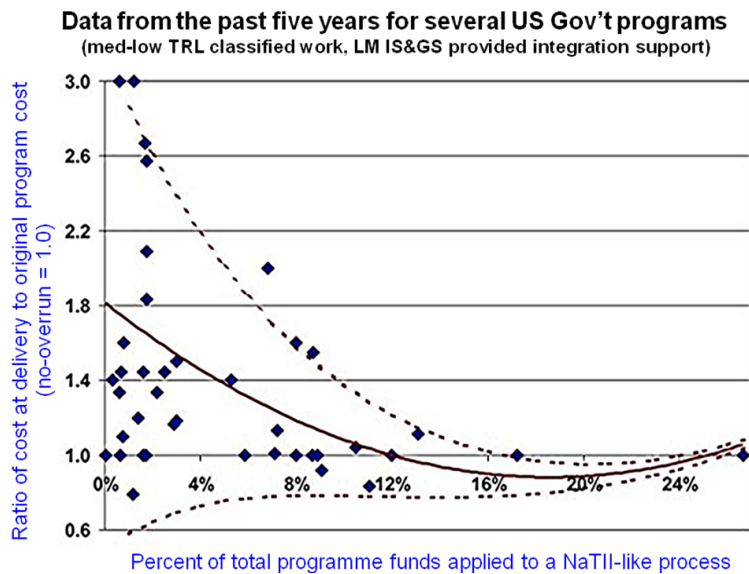
Figure 1: outline of the NaTII process – the square box represents the evaluation stage

The NaTII process is outlined in figure 1. It is mainly concerned with the evaluation stage, though some help and expert advice will be offered to those who take the project on through the implementation stage. In particular, NaTII can

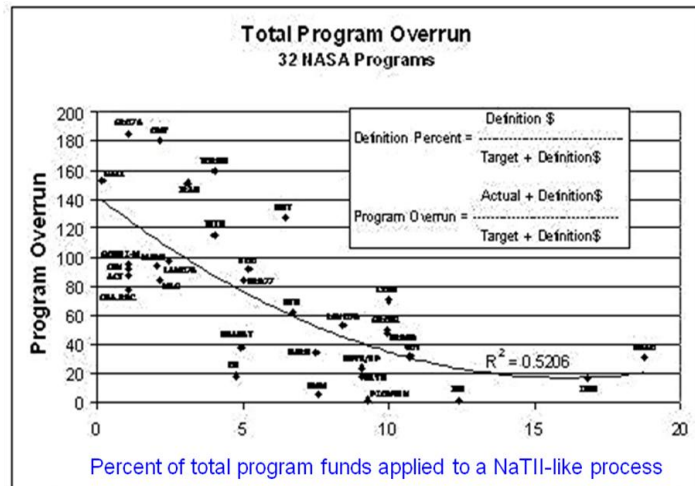
- Identify players/stakeholders, and technologies, to deliver a solution
- Work to transfer the knowledge gathered during the evaluation stage to the delivery parties
- Help to influence the release of government and commercial data where this can be shown to provide value in a solution
- Assist with security-related issues that may occur with release of data

NaTII is particularly relevant for those more innovative projects where success means high returns and where there is an element of risk. Studies have shown that front-end investment in these types of complex programmes is directly related to time and cost overruns and to the level of technology readiness level. Figure 2 highlights the inverse relationship between investment in up-front risk reduction activities (ie the kind of process NaTII will use) and cost overruns. Both graphs show that projects with little or no up-front investment tend to overrun on costs.

With a minimum of 10-15% investment in up-front integration there is a dramatic reduction in programme cost overruns, and dramatic whole-life cost savings. In general projects that are rated on the TRL scale of 6 to 8 generally benefit from up-front investment of around 7-10% of the project cost. In the case of NaTII projects we are more likely to be dealing with projects in the TRL scale between 2 and 6, where a up-front investment of 10-15% has the potential to greatly reduce the risk of cost overruns. The benefit of this in monetary terms will be carefully assessed and evaluated over the initial 3 year period of NaTII's operation.



Data from NASA's 1980s programs (med-lowTRL)



Source Werner Gruhl  
NASA Comptroller's Office

Figure 2: benefits of up-front investment

## 4 Data

The combination of data belonging to different organisations has the potential to create applications of considerable value, but this is often not apparent without the opportunity to experiment. NaTII will provide facilities for experimentation, and so help to persuade owners of data to be more open to collaboration and undertake riskier projects that may involve high levels of innovation. The exploitation of large quantities of data is in its infancy. The development of new services will stimulate the gathering of new types of data and the scrutiny thus generated will lead to improvement in the quality of the data. To discover what services are possible, and what is their value, will require data owners to get together to experiment in a safe environment that offers the opportunity to assess proportionate risks.

Nowadays, issues of security, privacy and consent monitoring are of prime importance and NaTII must



pay proper attention to these. Recent events have led Richard Thomas, Information Commissioner, to comment:

Privacy matters more than ever before, especially as so much of our personal information is now collected and shared. Public trust and confidence must be earned through tighter security and other data protection safeguards. Retaining trust and confidence also relies on organisations not collecting or sharing excessive information in the first place. I will expect government and other organisations to carry out rigorous privacy impact assessments before new large collections of data are stored or shared.

NaTII will have to give help with all these issues. It will not be a data repository. It will be influenced by the model of how to handle data that has been provided by the DfT-funded study of a National Transport Data Framework (NTDF). The NTDF has been conceived as a repository for information about the data rather than a store of the data themselves. All data should be left in the care of their owners, who are responsible for keeping the data up to date and as accurate as possible. In particular, the data quality issue is real, it is difficult and it can be improved significantly with some effort and made even better with more intense effort. Maintaining the quality must be the responsibility of the owner of the data, which must remain in a location under his control, rather than being transferred into a central server. Exposing data through the creation of services that use them puts pressure on owners to improve them. And aggregation of data provides cross-checks that can and should be used to improve data quality.

Data owners must be able to lay down their own rules for access, giving different privileges to individuals or classes of individuals with the confidence that they will be adhered to. Conventional approaches to access to large data sets rely on rigid classifications of the data. They are designed to suit a particular application and it is not easy to adapt them for other uses. A set of technologies, which is known as the semantic web, allows a rich description and consequently better search of diverse data sources. The semantic web means different things to different people, and there is no unanimity on whether it is the correct approach. But what it seeks to do is easily allow information to be tracked down, and then subjected to large-scale linking and aggregation with other, very different, information. See [http://en.wikipedia.org/wiki/Linked\\_Data](http://en.wikipedia.org/wiki/Linked_Data) and, for a specific application, <http://www.sciborg.org.uk>.

It is often valuable to create a “mashup”, a means of visualising a solution to aid the assessment of the project. This may need to mix real data with synthetically-generated data. The idea behind a synthetic data source is that it can mimic what could be achieved if particular types of data were available, and so help assess whether it is worth the time and expense of gathering real data. A second benefit of a synthetic environment is that events can be simulated faster than real-time. Where the creation of the mashup requires significant computing resources, it will usually be appropriate to use the grid approach, that is to distribute the number crunching over a number of computers

Research still needs doing on how best to handle data. Some of it is outlined in appendix A.

## 5 Who is NaTII for?

NaTII will serve

- Central government
- Public bodies, eg Highways agency, Ordnance Survey
- Local authorities, Regional Development Agencies
- Transport operators
- Industry and commerce: both large organisations and SMEs
- The general public
- The environment

No single data collector or user, government departments included, can reliably predict how data may be used when combined with data from other sources. The collaboration that NaTII is designed to foster will be essential. NaTII will

- (1) Draw in government and other publicly-funded organisations to collaborate in sharing their data so as to develop valuable new services for the public good;
  - (2) Encourage commercial organisations to collaborate in sharing their data so as to develop valuable new services for the public good;
  - (3) Help commercial organisations discover what services might be commercially profitable and develop them as efficiently and rapidly as possible;
- and it will bring commercial and government organisations together to fulfil each of these.

NaTII's main function will be to perform rapid but intensive evaluation of possible projects to determine whether they are worth taking forward to implementation. In carrying out this function, the NaTII service has the potential to benefit both the project's sponsors, its users, and wider society.

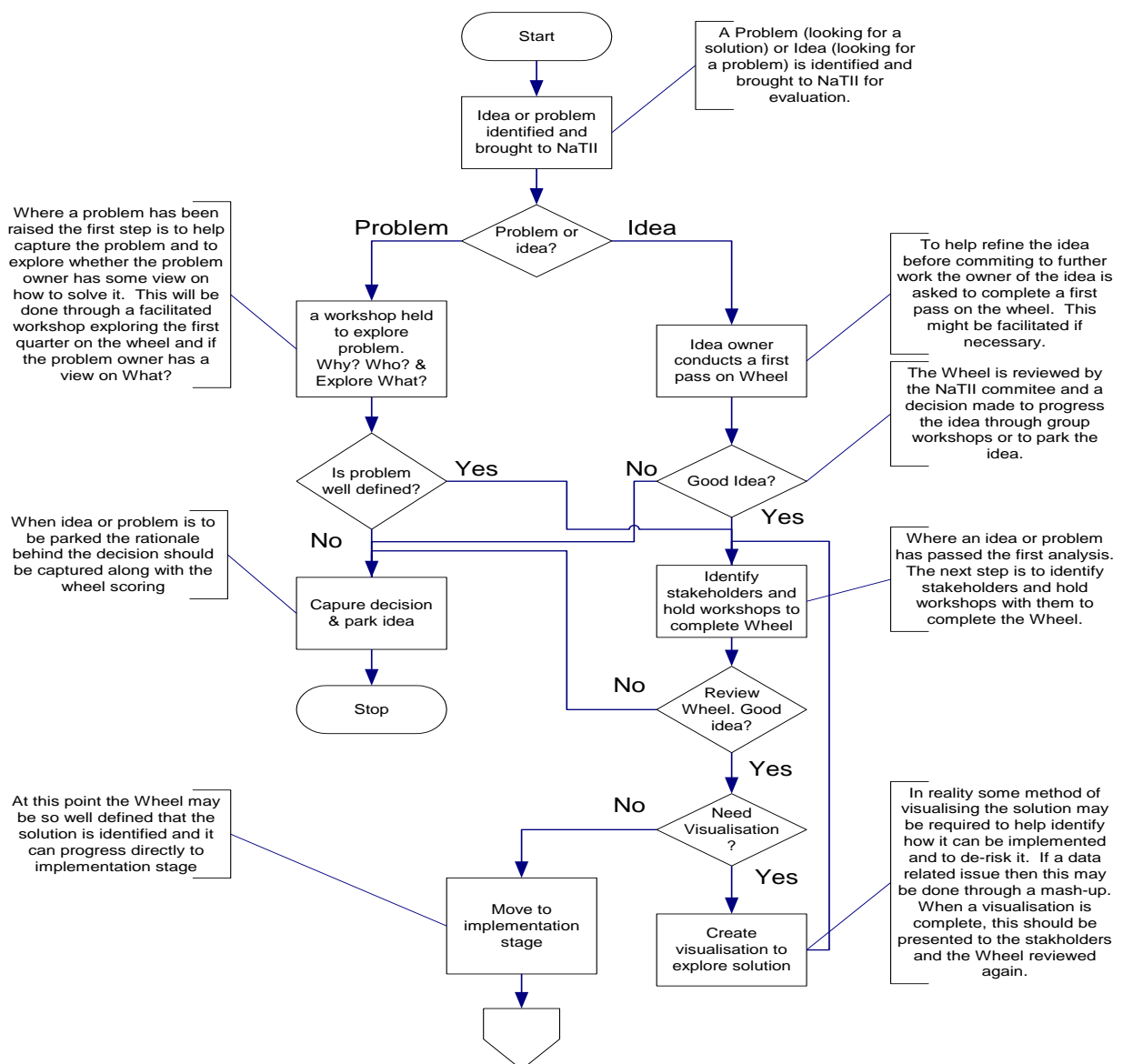


Figure 3: schematic outline of the NaTII evaluation process

## 6 Evaluation

NaTII's main purpose is to help decide the function and purpose of a project, and assess its viability, through a systematic evaluation stag, which has three basic components:

scoping

modelling

assessing.

The process is highly iterative. Each component is continually improved and resulting changes are then fed back so as to modify the work that has gone before. A schematic outline is shown in figure 3.

The **scoping** component involves workshops, and more private meetings, to identify in detail what the project is, who it is for, and when it needs to be implemented. It also seeks to identify other possible applications of the work. The **modelling** component mocks up the project in the laboratory. It uses real data as far as possible, and assesses whether their quality is sufficient for the application; where suitable data are not available it synthesises them. The **assessing** component repeatedly asks whether enough information has been gathered to decide whether the project is viable, so that at each stage a decision is made whether work on the project should continue.

For evaluating new ideas and deciding which one is worthy of further work, many of today's processes, although thorough, can be difficult to implement. In formulating new ideas, gaining the attention span of key stakeholders can be difficult. A process that can optimise their time and produce a quick result is ideal. It is aimed at creating a rapid process suited to evaluating ideas from all industry sectors and creating an action list for further work. It focuses on:

- visual and simple
- the fact that the first 20% of effort gets 80% of the result
- fast, short, focused, activities suited to a brainstorming workshop
- ball-park estimates for market size rather than requiring detailed market analysis.  
This can be done properly if the idea is deemed suitable to take forward
- Identifying the closest business model for the new idea
- producing a technology readiness assessment for any technology that is involved
- scoring the new idea and creating a visual method of assessing the result
- creating an action list to drive the idea forward

To support the go/no-go decisions an innovation process has been created. This consists of a tool aimed at achieving a rapid evaluation of ideas and the assessment of viability. It focuses on 8 key questions, taken in order:

Why? Who? What? How? Model? Players? Cost? Value?

The answer to each of the questions is then marked out of 10. The idea is to start in question 1 and work through the questions in order, gathering data through the brainstorms and workshops. Once a full set of scores has been completed, the exercise needs to be reviewed and the scores revised based on the results of the later ones. A low score can sometimes reflect lack of knowledge rather than a poor assessment. As an example, figure 4 shows our scoring for "My Journey", arranged around a wheel for easier visualisation. The low score on "cost" purely reflects our lack of information.

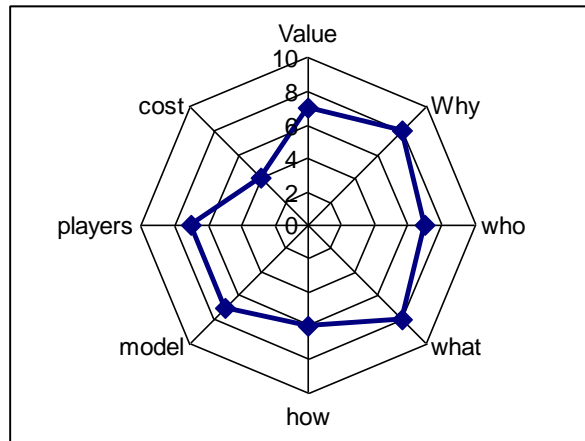


Figure 4: evaluation scores for “My Journey” arranged around a wheel

The key to a good innovation process is a method of ensuring that the wrong ideas are parked quickly, thus focusing valuable resources on the right ones. Inevitably, the scores assigned to each question will to some extent be subjective. However, those who participate in the scoring process bring in a variety of expertise and experience, and this should lend some confidence to the process. Ideally the tool would be used in facilitated workshops where key stakeholders can be brought together. The output from the workshop would be the scorings of the idea and an action list defining what should be done next.

## 7 Expertise

The NaTII team must have significant experience in leading major complex initiatives and research programmes. NaTII will offer breadth of expertise, with the ability to see the wide picture. The strength of NaTII will lie in the quality of its skill base, both in the core team and the pools of expertise.

NaTII will need to have sufficient expertise to help to

- Encourage different organisations to get together to experiment in a neutral environment
- Minimise the time needed to assess whether a project is worth developing
- Help to identify key performance indicators and success criteria
- Help develop synthetic environments to model ideas and so allow experiment to identify and prototype new services
- Help create synthetic data to test the quality of integrated data and determine whether it is worth gathering more or different sets of real data
- Help persuade owners of data to expose them in more useful ways
- Promote the gathering of improved data by demonstrating their value and exposing their existing defects
- Supply experience on such issues as data quality, developments in technology, the needs of the disabled, the environment.
- Help with underlying data brokering and security requirements that are needed to manage access to multiple databases with differing policies
- Broker data sharing between public and commercial organisations, so helping to manage access to multiple databases with differing policies, protocols, formats and operating systems
- Integrate a number of commercial and governmental stakeholders in an end-to-end process – this is unlikely to be achievable by a single commercial or government entity, and experience with large-scale systems integration will be valuable
- Involve end-users in the development of the process and the business plan, and the environment to allow exploratory changes to the process to investigate “what-if” type scenarios;
- Test how a new service could be delivered, to show how the service could be turned into a viable product;
- Help develop a business case to allow go/no-go decisions to be made on the idea;
- Hold workshops to develop the mashup and business plans.

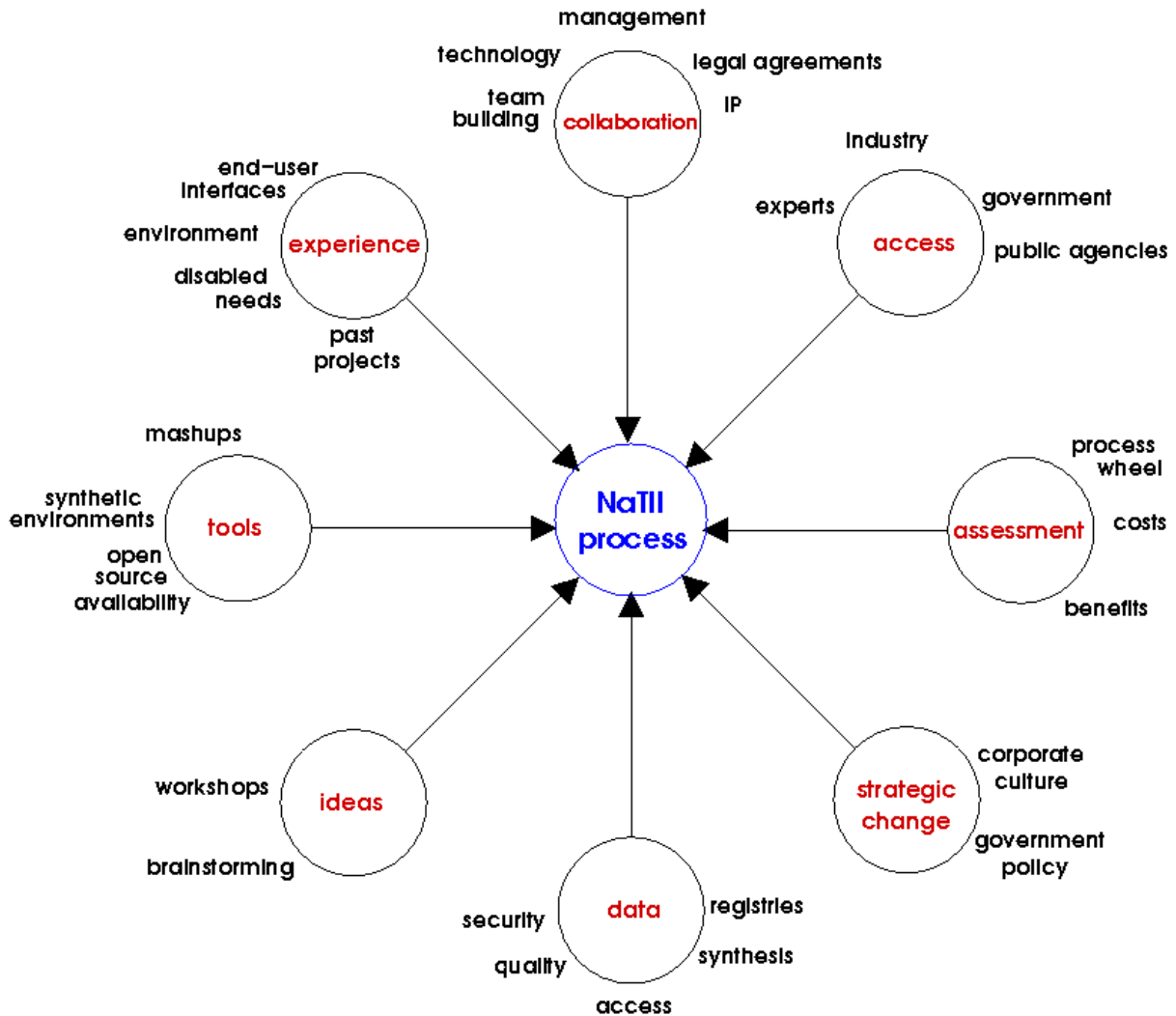


Figure 5: map of NaTII expertise

Each of the NaTII team will be expected to offer help in several of the following areas:

1. Data: security, quality, brokering access, synthesising, assessing quality
2. Ideas: through workshops, brainstorming
3. Tools: mashups, synthetic environments, open source availability
4. Experience: with technology, end-user interfaces, the environment, disabled needs, portfolio of past projects
5. Collaboration: team building, management, legal agreements, intellectual property
6. Access: to experts, industry, government, public agencies
7. Assessment: evaluation, costs, benefits
8. Strategic change: corporate culture, government policy

See figure 5. An important point is that NaTII does not aim to offer deep expertise; if that is needed it will have to be sought from outside and NaTII will help locate it. NaTII will advise on how to use the expertise. Some of this relates to technical issues such as understanding errors in data and their impact. There will also be help to develop business models to run a collaboration. There will need to be continuing development of tools and addition of new ones. Partly, this will happen through putting more and more projects through NaTII, but it needs some permanent staff to provide organisation and oversight. The various pools of expertise will grow as more and more projects are subjected to the NaTII process. Some of the issues and necessary tools for each will be application-specific, but they

will also leave behind an ever-increasing generic resource. NaTII will also need access to outside advice, partly to help it keep abreast with developments in technology but also to bring in fresh ideas. There will be panels with wide expertise and experience to brainstorm what could be of value and how best to achieve it. They will also help with ideas on how to obtain direct income for NaTII.

There will have to be staff who go out and explain to organisations why it is valuable to consider the sort of collaboration that NaTII will foster. The case for collaboration will become steadily more obvious as NaTII builds up a portfolio of successful projects. A central aim of NaTII will be to get people to do what they did not know they wanted to do. NaTII will use its ever-increasing experience with its participants to propagate innovation through the industry supply chain of each project.

## 8 “My Journey”

My Journey was the particular application that we used to demonstrate the feasibility of the NaTII process and to learn how the processes and tools could be used in practice. BAA Stansted are the My Journey end user who would benefit from the service provided. In order to maximise the BAA Stansted asset, BAA would like to see an increase in the throughput of travellers at the airport. Stansted also already has one of the highest numbers of public transport users in the UK. They would like to maintain this lead and increase asset usage, through the provision of travel-related services. The aim of My Journey is to increase BAA Stansted incoming passenger use of rail plus coach by 10% from the 4th quarter 2006 baseline, as well as increase the number of people purchasing tickets remotely. In addition, it should allow BAA Stansted to improve its score in its Quality Service Measurement surveys and reduce passenger loiter time from the arrivals baggage hall to surface transport.

The work on My Journey led us to

- Demonstrate the principles of real-time information service incubation using advanced web technology
- Show by using a real transport problem how a useful information service could be developed in a very short time – weeks
- Show how personal information could be kept secure in support of personalised services
- Investigate the business case for a commercial service of this type
- Identify outstanding research and innovation issues
- Identify the market for such services

We spent some 4 months of our pilot study examining the feasibility of My Journey, and devised a scheme that uses the internet to produce real-time journey planning delivered to passengers through their mobile phones. We were told that, normally, BAA would have needed 6 to 9 months to select and brief consultants, and would have anticipated that it would then take them 6 months to do the work. Also, the consultants would not have assembled the mixture of high-level experience and expertise that we were able to apply.

BAA made these comments on the My Journey project:

- One of the initial reasons for getting involved with NaTII was the quality and calibre of the consortium partners. It was proven that deliverable operable systems can be conceived from risk technologies, that is those at low technology readiness level.
- This proved successful in that we were able to form pools of excellence very quickly and were able to bring high skills and diverse expertise around a table.
- The influence of Cambridge was highly beneficial
- Focus was on the end user.
- The process was driven against a time scale and key dates were met with deliverables, very important when looking at new business propositions.

We understand from BAA that the benefits included

- An outcome in effectively a 4/6-week period – a saving of some 12 months over BAA's normal business process
- Verification of a complex process to agreed dates
- A high level of confidence in achieving the development budget
- Effective reduction of front-end cost, and a range of expertise and knowledge and consultants that you just would not normally get.

## 9 Project identification

The barriers that have been encountered so far with collaboration among different organisations, for example those experienced with the *Data Grand Challenge* initiative (see section 11), suggest that, at least initially, an important function for NaTII will be to go out and look for projects to put through its process. As it builds up a portfolio of successful projects, and knowledge of NaTII's value spreads, we envisage that this will stimulate organisations to present their own projects, though they may need help in recruiting appropriate partners into a collaboration. So a very important function for NaTII will be that of putting in place projects that appear to have the potential to become valuable business propositions or produce worthwhile services for the public good. Although one aim of NaTII is to reduce risk, it cannot be completely eliminated and those who propose projects must be willing to take calculated risks.

Some projects will be driven by a perceived need for a new service, others by the existence of some capability: *I have these data, how can I extract value from them?*

With any proposed project, the first task will be to assess whether to start to put it through the NaTII process. Of course this means partly judging its apparent intrinsic worth, and partly deciding its priority among other competing proposals. We suggest that to begin with, and until pressure on NaTII's resources builds up, it will be sufficient for the Director of NaTII to take advice from the NaTII staff, since they will have the wide range of expertise detailed in section 7. In some sense, the decision to initiate the NaTII process is not a major one because, as we explain in section 6, the process is iterative and so should achieve a rapid initial assessment and projects that are judged to have relatively little value will not be taken very far in the process.

Issues of transparency can be taken care of by the need for the Director to report to the NaTII programme management board, which includes representation from the funding body. However, before long it will probably be necessary to set up a dedicated project advisory group of expert outsiders who are perceived to be neutral. In either case we suggest that, while the Director should normally take the advice offered, he or she should have the final responsibility for deciding which projects to take forward, because it is the Director who is charged with making NaTII a success and is responsible to the funding body for ensuring that. This arrangement has been followed by other publicly-funded projects.

## 10 Examples of projects

Three possible projects were identified at a workshop in December 2006, which brought together two dozen organisations from industry, commerce and the public sector. They identified three projects on which collaboration would be of value and which were further developed by the authors of this report:

- My Journey: information for incoming airport passengers (see section 8)
- My Travel Footprint: web based provision of traveller carbon footprint calculations
- My Event: mobile information feed to major event attendees

These have been developed further in our pilot study. For the pilot study, My Journey was the most advanced application and it was chosen as the one to demonstrate. The other two applications are still being progressed.

We have been discussing various other projects. Among them are

- 5-dimensional interactive modelling simulator

- 21<sup>st</sup> century transport infrastructure in the digital age.

The first of these has attracted interest from the Ordnance Survey, and the second is a suggestion from the Mobile Data Association.

### **My Travel Footprint**

Organisations are beginning to take an interest in trying to minimise their staff's carbon emission in travel to, from and at work. Several tools are being created to calculate carbon emissions. A workshop on 18 February 2008 brought together experts from the DfT, the Energy Saving Trust, the Carbon Trust, the Open University, and a number of large companies, to tease out how to take these in imaginative directions, make them so that staff are willing to cooperate, and persuade employers that this is a worthwhile exercise. Various issues that are rehearsed in this report were brought up by the participants, for example the need for a safe environment to share data, the need for data compliance, and the problem of data quality – on the whole, only people who are proud of their travel habits will respond to questionnaires.

### **My Event**

The Department for Culture Media and Sport is concerned that the 5 million additional people who will be in London for the Olympics shall not all try to go to the same place at the same time. So they are interested in extending My Journey so as to use mobile phones to supply information about the many things that will be going on in London at the time and how to get to them. As experience has shown that large-scale IT projects usually run into trouble, we have decided that it is necessary first to mount a pilot. AEG Worldwide, who operate The O2 dome, have expressed interest in this.

### **5-dimensional interactive modelling simulator**

This project would extend the work on NaTII by creating an Interactive Modelling Simulator to help improve the understanding and behaviour of new urban regeneration developments and other major complex new projects. The essential feature will be a 3-dimensional virtual environment in which one can test and simulate real time information and data. The model will be interactive with time and cost; in this sense the simulator will be 5-dimensional.

### **21<sup>st</sup> century transport infrastructure in the digital age**

The Mobile Data Association has suggested an ambitious project that, among other things, addresses connected cars and connected roads so as to improve UK infrastructure and competitiveness. There is already a great deal of activity in this area but, because of its neutral character, NaTII could lead policy initiatives and agreeing of standards, facilitate data sharing and collation, for example safety statistics, and generally be a trusted agency.

Further discussions have started on some new projects. These include:

- Expansion of MyJourney with specific users: Network Rail, Go-Ahead
- Transport information aggregation: investigate how rail information could be shared across providers to improve efficiency and effectiveness
- The O2 dome: event crisis management

An early task for NaTII should be to undertake a trawl through the DfT for other projects.

## **11 Barriers**

Various barriers were identified in the series of meetings the DfT organised to take forward the *Data Grand Challenge*:

- Data Sources - what data is available; where and how is it stored?
- Data Access - how is data made more accessible to legitimate users?



- Data Use - what limits are placed on the uses to which data can be put?
- Data Confidentiality - how can personal data be protected or anonymised?
- Data Security - how secure is the data from unapproved access?

Others have been identified during our pilot study. By adopting new approaches, the barriers that NaTII will seek to address include:

- Each member organisation in a collaboration brings in a set of lawyers and, with each lawyer feeling the need to contribute to the negotiation, it takes much time and patience to put agreements in place.
- Organisations generally develop their own conditions for allowing others access to their data and so a large number of separate negotiations is needed.
- Creating new services may well suggest further services, but until these have been identified they cannot be valued.
- New services may take a long time to perfect and to be recognised as being valuable, and so it may be some time before there is a return on investment.
- It is difficult to put a financial value on services that are put in place for the public good.

NaTII will help with alternative contractual arrangements. It will establish different data access protocols and new and accelerated evaluation processes to verify value.

We have consulted experts on valuing services for the public good, both within DfT and in our own organizations. Part of the problem is that one cannot foresee all the benefits, many of which may be indirect or accrue in the longer term. Although we understand that a number of techniques have been devised to value these, a degree of subjective judgment is inevitably required, which people are reluctant to provide because of the risk.

## 12 Organisation

Our recent intervention with industry has exposed a number of key issues. NaTII is liked for its neutrality of brand, the quality and integrity of the people drawn from major industry and the association with leading academia. It is the quality of this mix of skilled people with large complex project experience that acts as a key attractor to organisations and individuals. It is therefore important that NaTII is structured to enable these qualities to be maintained and refreshed. These considerations have led us to suggest the structure:

A collaborative unit is formed with each organisation seconding key players into NaTII. The collaborative unit selects a director, who is responsible to the principal funder. Each organisation through its parent company enters into joint and several arrangements with NaTII in order to supply it resources and services

The Director should be perceived to be neutral, that is, unlikely to derive commercial benefit from any of the projects that might be expected to go through NaTII, and have experience of relating to industry, government departments and academia. He or she should carry some prestige, either in their own right or through the organisation to which they belong. The Director should be supported by a core of staff covering the expertise listed in section 7. In order to provide continuity, most of these, like the Director, should be semi-permanent. However, in order to bring in fresh ideas and expertise, others should be seconded to NaTII on short-term contracts, one year say. As we have explained, the expertise of the staff should be broad rather than necessarily deep, and so one person will cover several of the items listed in figure 5.

NaTII should include a core team made up of individuals who would be committed full time to the NaTII programme. They would be part of a programme management board, which would also include a representative of the principal funder. The Director would delegate key duties to the core team, to be responsible for managing the organisation and delivery of such duties. Typical of such roles would be:

- Programme Management
- Systems Integration
- Thought Leadership management
- Business development
- Pools of Expertise
- Information and security governance
- Tools development
- Business model evaluation
- Data quality
- Data compliance

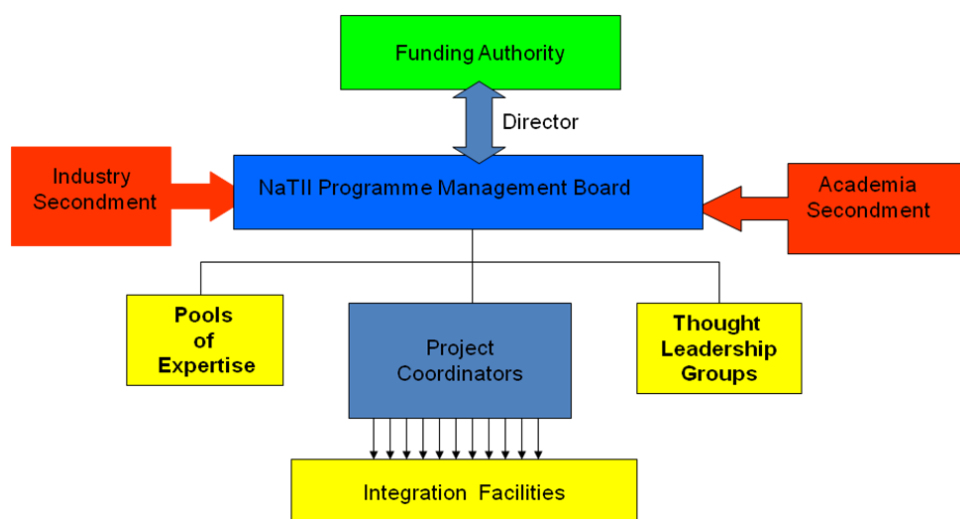


Figure 6: organisation of NaTII

In order to manage the different programmes, project managers will be brought in. These would be full-time roles for the duration of the various projects. These could vary in length from 1 to 3 months. In order to ensure that NaTII meets the stringent requirements of data protection and data security it should have a dedicated officer responsible for continuity throughout the NaTII process.

## 13 Costs

At least initially, NaTII will need to be funded by government. This will help it to be perceived to be a neutral territory, and it will also enable it to become effective and demonstrate its value, all of which are needed to encourage organisations to engage with it. However, we envisage that the government funding can be reduced after an initial three-year period. Some projects that are put through NaTII will be for the public good and may not produce any direct revenue, but for commercial projects that extend into the development and deliverable stage, opportunity exists for NaTII to earn fees and royalties, or participate in the end users' revenues streams.

Nearly all of the cost of NaTII will be people, though some further tool development will be needed. We set out in tables 1 and 2 below an approximate estimate of the cost to set up and run NaTII over a 3-year period. The headings have been established based on experience derived from our pilot study. The budget is based on parties to NaTII having access to integration facilities; it does not provide for the building of a dedicated facility. The NaTII environment will be set up in year 1 and hence only a small number of projects will run through the facility in the first year. The project costs are a rough

estimate as some of the projects may not make it past the first assessment. The project management costs also cover proactive events held to identify new projects and grow the expert network. As NaTII becomes established, the NaTII core team costs may possibly reduce, to be replaced by more project-specific activities.

	<b>Cost Item</b>	<b>Cost (£'000)</b>
Setup	Mobilisation and setup	200
<b>Total Set up</b>		<b>200</b>
Year 1	NaTII core team	833
	Pools of expertise	167
	Infrastructure support services	167
	Projects year 1	550
<b>Total Year 1</b>		<b>1,717</b>
Year 2	NaTII core team	833
	Pools of expertise	166
	Infrastructure support services	167
	Projects year 2	1,100
<b>Total Year 2</b>		<b>2,266</b>
Year 3	NaTII core team	833
	Pools of expertise	167
	Infrastructure support services	167
	Projects year 3	1,650
<b>Total Year 3</b>		<b>2,817</b>

Table 1: estimated cost of setting up and running NaTII

<b>Total Costs</b>	<b>Cost (£'000)</b>
Set up	200
Year 1	1,717
Year 2	2,266
Year 3	2,817
<b>Total (assuming 24 successful projects)</b>	<b>7,000</b>

Table 2: summary of cost

A description of each of the individual cost items is detailed below.

### ***Mobilisation and setup***

This refers to the establishment and mobilisation period which will be necessary to establish a NaTII environment. This is an essential period to ensure that the incubator operates efficiently and is equipped with the initial tool set. A right-first-time approach is required, particularly when dealing with government personnel and strategic people in business. The estimate is based on the internal costs paid by consortium members to get NaTII mobilised.

### ***NaTII core team***

The NaTII process demands the highest level of management and capability. The team is an operational one and will be a full-time commitment. The costs are non-profit but include admin and a contribution to general overheads. The budget allows for a team of 6, including the Director, all with allocated responsibilities, which include overseeing the programme, technology support, finance, networking, legal matters, and issues related to data, such as quality and compliance with regulations. More details are given in section 12.

### ***Pools of expertise***

Under the pilot programme several experts engaged with us without payment. In the full-blown NaTII we would expect much of the expertise to be paid for. The sum shown reflects the top quality and nature of the resources. It allows for establishing and running weekly engagements with the pool of expertise, necessary to the timescale allocated to the projects that will be subjected to NaTII.

Provision needs to be made to engage with academia and research institutions, to provide access to emerging sciences and technologies. It will give the end user access to key influencers and strategists to help thinking and direction, a key quality built in to NaTII. The budgets allow for ½ day engagements with experts, including visits to research centres

Thought leadership programmes are essential for integrating academia, government and industry. They are breeding forums for ideas and innovation. The budget caters for 6 per year. The costs include facilitators and facilities.

### ***Infrastructure support services***

It would be expected that consortium member organisations would be large enough to have infrastructure facilities such as computing at their disposal, so that activities can be distributed across consortium members as needs arise. The budget figure here represents a rental value for the facilities with additional cost for the adaptation of the facilities to meet the various needs of specific projects.

### ***Project Costs***

A dedicated manager who has quality skills in integration and facilitation will be assigned to each project. These are highly skilled specialists. NaTII will buy them where necessary, but in the main they will come from existing resources.

The budget represents the costs associated with developing the various architectures and running the various mashups. The majority of the cost relates to software integration, mockups and visualisation. Not all the projects put through NaTII will be judged to be viable, and because of the iterative nature of the evaluation process those that do not get through to the implementation stage will take varying times for the decision that they be parked. The costs are based on running 4 successful projects in year 1 followed by 8 in year 2 and 12 in year 3. An average of 92 man days is estimated for each project running through the full NaTII process, made up of 32 man days in meetings and workshops to prepare storyboards, report writing, expert input and mashup preparation, 40 man days for creating demos and visualisation programmes, and 20 man days for finalising reports and business plans .

## **14 Implementation**

Data and knowledge integration are central to government policy. To support this, there should be an accelerated procurement process, so that NaTII can begin to deliver results by early in 2009.

Written tenders will be largely replaced with presentations and interviews. It is essential to maintain momentum on this project. Organisations with the correct skill set and understanding of the strategy will have no problem in responding quickly in an innovative way. A review panel should be selected, with understanding of innovation, and there should be a 3-stage process:

- Call June 2008
- Responses September 2008
- Offer letter November 2008

NaTII should be up to speed with its processes and ideas forums so that its first project can begin by 1 January 2009

We propose the following high-level objectives for the NaTII programme:

- By 31 December 2009 4 projects will be required to be completed.

- Year 2 should deliver a minimum of 8 projects.
- Year 3 should deliver a minimum of 12 projects.

These figures are to be regarded as indicative. The time needed for a particular project will depend on its complexity, and for those that are judged not to be viable the iterative nature of the evaluation process may lead to this conclusion being reached either rapidly or only after a number of passes through the process. Several projects will be run concurrently and the number that can be handled will increase as NaTII gains experience.

The focus of the whole NaTII programme is based on achieving deliverables and measured outcomes. The intensity of the activity, if it is going to have impact, demands seeking out integrated projects that can carry over into the supply chain. From cross-transport activities we considered during the pilot, we are confident that a minimum of 24 projects can be delivered in the first three years. The budgets recognise that some projects will be discarded in the early stages, but there should be a significant number taken to implementation stage.

Specific deliverables and outcomes should be defined for each project that NaTII undertakes, and when completed each project should be evaluated against these deliverables and outcomes. This will allow NaTII to demonstrate project successes, learn from mistakes or failures, and measure its success.

As described in section 13, NaTII will have very limited if any fixed costs, given the majority of its costs are people-related and vary with the level of activity undertaken. Therefore potential funders can commit to funding NaTII for the initial 3 year period but funding need only be provided in stages, for example on an annual basis. Continued funding over the entire 3 year period can therefore be subject to conditions such as evidencing successful delivery of projects.

There should be milestone reviews with the funder on a monthly basis.

## 15 Acknowledgments

We have benefited from briefing sessions to learn about expected developments in technology, and other issues:

12 November 2007: Mike Short (O2 and the Mobile Data Association)

20 November 2007: Nigel Shadbolt, John Darlington and their colleagues at Southampton University

23 November 2007: David Cleevely, an on the digital economy

7 January 2008: Jerry Fishenden, Andrew Herbert, Peter Key, Ken Wood (Microsoft UK)

21 January 2008: Professor Andy Parker (Cambridge eScience Centre)

11 February 2008: Professor Sir David Williams and Dr Rufus Pollock (Cambridge University), Duncan Salmon and Paul Renney (Campbell Hooper), Roger Evans (Creative Learning Consultants)

18 February 2008: Mike Lynch (Autonomy)

In addition, we have had many meetings with Julia Gregory of BAA, James Brayshaw and colleagues from Ordnance Survey, Mark Hayes of the Cambridge eScience Centre, Guy Wolfenden of CitySpace, Rosa d'Alessandro and colleagues from AEG Worldwide, Tom Steinberg of My Society, and Mark O'Neill and colleagues from the DCMS.

We particularly acknowledge the support of Nick Illsley and Paul Drummond of Transport Direct, and John Sheridan of OPSI. Finally, this work would not have been possible without the active help of Professor Brian Collins, chief scientific adviser at the DfT, and his team.

## Appendix A: Research

There is economic and policy research to be done on the incentives to make data openly and freely available, recognising that there is more money to be made from higher-level services and applications than from the data themselves.

The way in which data are handled and made available for easy large-scale use is developing rapidly and it is an important area for discussion and research. Our thinking has been influenced by work that we commissioned from academics at Southampton University (see section 15). As part of their work for My Journey, they marked on Google maps the locations of all the ATMs in two London boroughs. We have also been shown how the Ordnance Survey, using its vast Digital National Framework database, can rapidly plot on one of its maps with high accuracy the locations of all the ATMs in any area of the UK.

Ordnance Survey data have to be paid for. Part of what NaTII can do is help provide evidence to enable a decision on whether making them freely available could be paid for from the greater tax take to the Treasury resulting from the resultant increase of economic activity. Cost benefit analysis is commonly used by Government to make decisions. For example, decisions on new transport links are normally made by comparing the costs to a model of the notional benefits to the economy in terms of reduced journey times and congestion. Research is needed to develop a similar model to quantify the benefits of shared data. This is more difficult to achieve for intangible services provided by software applications, but the very high valuations placed on internet-based businesses which have been in existence for only a few years indicates strongly that there is a very large potential increase in economic activity from data sharing. A cost-benefit study could investigate valuing data sharing activities on the basis of the number users expected for the services (this is often the basis of internet company valuations). The value of startup companies which might be generated could also be estimated. The commercial sector recognises the value of data, as evidenced by their attempts to capture information on their customer bases by various means, and the trade in such databases. Such companies could participate in research on the value of shared data.

On the other hand, the public perception of data sharing needs to be investigated. There is widespread suspicion of sharing of personal data, but significant support for the release of data seen as being of public benefit, such as the results of drugs trials. It is possible that the release of publicly funded, impersonal data, such as that held by the Ordnance Survey, would be generally welcomed as a fair return to the taxpayer. There are already calls for action in this direction, such as the "Free Our Data" campaign in the Guardian newspaper.

The Southampton approach is based on the techniques of the semantic web (see section 4). Semantic web technology facilitates faster optimization through faster reasoning. Combining and distributing human and machine decision making in My Journey, the Resource Description Framework (RDF) used by Lewisham and Camden boroughs facilitated faster correlation between precision locations of objects provided by the Ordnance Survey and supplemental information provided by local authorities and commercial websites. It is faster because many diverse groups can develop a representation of data, which can then be brought together in a mashup quickly and easily. This needs a lot of work but the work can be distributed across many people so it is not a burden on any one individual or organization. Nevertheless, there is a view that RDF is over-complicated, that semantic web techniques should be imposed on data collectors only when a well-defined purpose is in mind, and that it would not be a good use of resources to try to impose them more widely.

The NTDF study adopted a metadata warehouse approach which attempts to capture the benefits of RDF without imposing the burden on the data holders. RDF is used to manage a repository which contains a description of the data, and how to access it, rather than the data itself. It also contains information on the access rights of users. Users can treat the metadata like an index to find the data they require, and applications can be built to access and combine the data in any way required. In this way, the complexity of RDF is reserved for the tasks which really need it. Organisations holding data have no need to modify their current practices, or to adopt RDF. All they need to do is to maintain a correct description of their data, and of the access rights they have granted, in the central repository.

The semantic web approach emphasises the value of imposing a particular structure on archived data, and there are important issues related to data that have a much less well-defined structure. Current techniques for extracting structured data from unstructured text are unreliable and time-consuming to develop and maintain. Better tools are needed to aid the publication of structured data at source and to extract it from unstructured and free-text sources. For example, weblog and online news publishing systems now routinely and automatically produce RSS feeds for their content. This could be extended automatically to external data sources such as wikipedia. Reuters have recently developed the Open Calais system <http://www.opencalais.com> which will automatically extract place names, people, companies and events from raw HTML. There is a large expert system behind this and its results so far are limited, but it should be possible to extend the ideas so as to recognise wider classes of objects and automatically insert links to relevant web resources and data.

More generally, research is needed on how better to control access to data and how to manage access to different layers for different people. Use of real-time data is in its infancy, and raises new issues of security and access control. Then, at a more commercial level, there is the matter of how best to identify and share out transaction costs, so as to organise and broker micro-payments in a way that most fairly provides rewards for sharing data.

Security is another vital area where research is needed. Architectures exist that allow secure data sharing, but using them more widely will require government standardisation and accreditation. Note, however, that seeking to impose data standards is not necessarily wise: it is often better to allow them to develop under the influence of some external requirement, such as the desire to display information on Google maps.

## Appendix B: the authors of this report

**Peter Landshoff** is Professor Emeritus of Mathematical Physics in the University of Cambridge. He was one of the leaders of a £60 million pound building project in Cambridge, which became the British Construction Industry's Major Project of the Year, 2003. Working for the Cambridge-MIT Institute, he learnt how to engage with industry and government, on projects that include the construction industry, transport, and the design of urban communities. In Cambridge, he has been closely involved in the creation, and then the running, of a number of projects. Among them are the Newton Institute, the Cambridge eScience Centre, the National Institute for Environmental eScience, the Cambridge Computational Biology Institute, the Centre for Quantum Computation, and the Millennium Mathematics Project.

**Michael Simmons** works as development manager in the University of Cambridge Computational Science Centre, helping identify and support new and existing projects (See <http://www.escience.cam.ac.uk/projects>). He has been co-ordinator for several multi-stakeholder projects. He has a background in office computer networking and university and industry relationship development including working with Sun Microsystems in a dual university and industry role to support collaborative projects. He has degrees in computer science and history.

**Dr Rashmi Joshi** has over 8 years experience in the field of data analysis, incorporating a broad spectrum from data extraction to validation, analysis (conventional and innovative) and statistical consultancy. She has worked extensively with conventional modelling techniques (regression, stochastic) and is capable of formulating and developing innovative predictive models using cutting-edge techniques, based on a dynamic class of mathematical models known as artificial neural networks. She has created an artificial neural network for modelling survival time data which was successfully applied to NHS data sets in order to devise and quantify strategic policies. She has acted as chief statistical advisor within the NHS, and is responsible for the annual consultant appraisal within the NHS Trust, from development of a system for surveillance to standard monitoring processes.

**Charles Gill (Chuck)** is an experienced programme manager and engineering technologist. His specialised skills are in large information collection and processing systems associated with the US and UK intelligence communities. During his 30 year career, he was the chief of the joint UK-US

facility at RAF Menwith Hill, North Yorkshire, where he successfully led the completion of a dozen, multi-million dollar projects including new Information Technology systems and improvements to the station's facilities. Before that, he led a world wide monitoring system linked to the International Atomic Energy Agency. He has managed a multibillion dollar, worldwide network of intelligence stations, run a state-of-the-art nuclear forensics laboratory and was a combat aircrew member in the US Air Force with over 2000 hours on ten different aircraft. He has worked for Lockheed-Martin as a systems engineer for less than a year, but he has developed and delivered effective alliances between diverse US and UK government agencies as well as their associate contractor base.

**Colin Waugh** joined Thales Research and Technology (UK) Ltd (then Racal Research Ltd) in 1984. He has worked on and managed many projects in RF and digital engineering during his time at Thales including innovative work on HF and VHF radios, satellite systems and digital techniques for radio receivers. Before moving to his current role as Business Development Manager, he was the engineering resource manager at TRT (UK) and lab head for one of the application areas. He has many years of experience in process development and innovation. In his current role, he manages the marketing activities of TRT(UK) and is responsible for developing new areas for the application of TRT(UK)'s technology expertise. He has an honours degree in Electronic Engineering, a Certificate in Management from Reading University and is a Member of the Institution of Engineering and Technology and a Chartered Engineer.

**John Patman** is a business strategist and engineering technologist with a specialised skill in contract strategy, processes, front-end customer consultancy and the development of customer strategies for improving business performance. Much of this work was developed as a director in the Taylor Woodrow Group until 2001, when John started up his own business Syzygy-cbs Ltd. He is well experienced in project management, property development, offshore engineering, construction and management contracting both in the UK, North America and Europe and Far East. Among his many projects have been the first of the North Sea oil rigs, the Channel Tunnel; the computerized ticketing system for LUL; BAA Terminal 4 and Heathrow Rail Express.