

Business Information In Government

Summary Report

Introduction

This is a summary of the final report of the Invest to Save Budget funded project called the Business Information In Government (BIIG) and its concluding Comprehensive Business Directory (CBD) study. The work began to address a need for better business identification and to support future joined-up government services.

The project has been a research process exploring the general needs of businesses and government services for unique identification and, based upon these findings, examined ways that public services could address these needs.

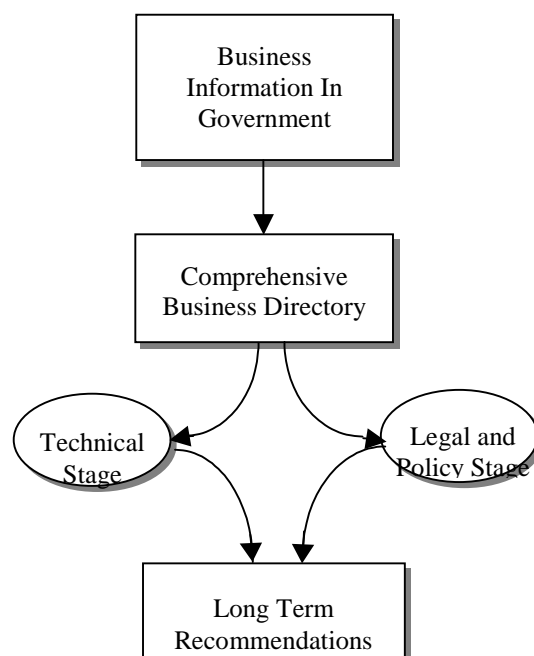
To ensure that the research covered a spectrum of government interactions with business the following departments agreed to participate in the research with the Office for National Statistics providing project management and support: Companies House, Inland Revenue, HM Customs & Excise, Department of Trade and Industry, Office for National Statistics.

Collectively these departments provided an extensive knowledge of the key areas of research, namely:

- Business Referencing
- Business Registration
- Business Data Management

Project Conduct

The project ran between April 1999 and August 2001 with the work split in two phases. At the end of the first phase (Business Consultation and Information Flow Modelling) the results were presented to the Information Age Champions which led to a further step that resulted in the paper *Electronic Service Delivery for Business*. Phase 1 and the subsequent work confirmed that a unique number for all government purposes is not possible and that the management of a free-standing register would create weaknesses that would prevent it being used as a replacement for critical departmental functions. Phase 2 focused upon the solution proposed in Phase 1 for a multi-functional product that could build upon the existing information sources both in the public and private sector (later referred to as a Virtual Data Hub). The proposal raised a number of questions in regard to the legal position of data holdings and the technology configuration that would be necessary.



Summary of Findings - Phase I

A number of significant conclusions have emerged from the Phase I work. Some of these run counter to the conventional view of the way forward and have important implications for the future direction of policy concerning business information in Government.

The main findings are set out below:

The need for a holistic view

It is clear that any consideration of business registration and referencing has wide implications for business and Government, and that any business referencing scheme designed in isolation will be unlikely to succeed. Unfortunately, many of the propositions so far advanced largely fail this acid test.

In order to succeed a regime for managing business information in Government needs to consider the broader agenda of 'Modernising Government' and how better joined-up service provision can be achieved. At a micro level, the impact on procedures, forms and current IT systems also needs to be considered.

Developing a holistic view also requires scenarios for the longer-term shape of Government to be factored in. Any solution to the business registration problem must, in any case, avoid constraining future service delivery options.

Business referencing

A better means to identify and reference business across Departments would be beneficial. However, this can be achieved without the imposition of a Single Business Number (SBN). Indeed, to do so would create severe practical difficulties for limited benefits. Business numbering schema are embedded (often deeply) into departmental processes and systems and hence hard to change. Also, it is difficult, perhaps even impossible, to define a numbering scheme applicable to all the different ways in which a business might deal with government both now and in the future. The benefits of a SBN to business appear slight.

A better approach will be to develop instant recognition of a business and allow flexibility in how identification is achieved that best suits the role/purpose that business assumes for each dealing with Government.

Business registration

Although registration is a key event in the business life-cycle, there is no single registration function across all departments. Rather, businesses register for a variety of purposes and in a variety of roles (e.g. employer, VAT trader or polluter). Each of these roles has different information characteristics.

Therefore, the concept of a 'single business registration' process is not valid. In preference, there should be greater harmonisation between separate registration systems with a view to linking where this adds value.

Storing and sharing of data

A comprehensive common or 'core' dataset across all departments does not exist. Departments collect data according to their own (narrower) needs. Even within common data such as post code and address, there is little commonality. Previous attempts to define a

common data set have not been successful and therefore, the answer lies in achieving a collective view through controlled data matching and sharing (subject to legal permissions).

The matching regime would build upon the high quality registration numbering systems such as Companies House and VAT - and newly harmonised registers extended to data sets such as name and post code for further verification.

Rationalisation of business interactions

Government departments sometimes request similar data from a business several times. Rationalisation of such interactions requires an 'intelligent gateway' able to distribute common data to the relevant government departments. It also requires rationalised departmental processes to deal with the rationalised common data. The immediate scope is limited by the difficulty of defining common data and should focus on registration of information in some key business role(s).

Summary of Findings - Phase II Legal and Policy

“Generally, the legality principle applicable in all the Member States requires that all administrative acting must be performed according to law. Government agencies are only competent to take action to the extent that they are empowered thereto by the law. No legal action may be taken without legal basis. Public authorities need to be vested with expressed or implied power in order to co-operate with each other.”

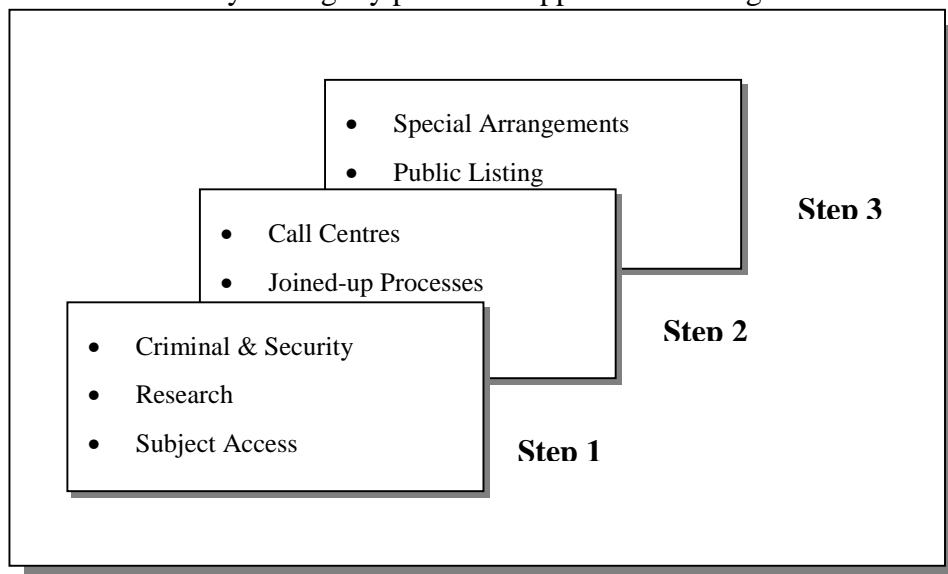
<http://europa.eu.int/ISPO/ida/techlib/legasp.doc>

Currently permissible developments

Within the current legal scope, it is possible to use an access facility for a limited number of purposes:

- **Criminal and security**
- **Research**
- **Subject Access**
- **Data already in the public domain**
- **Within specific existing legal gateways**

Further developments would require legal change. Such developments would need to be incremental and dictated by the legally permitted opportunities for government closer



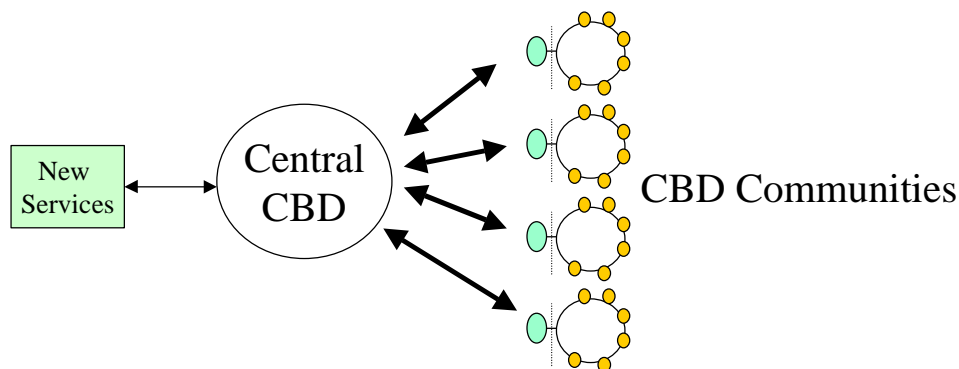
working. This can be represented as a three-step development with each step serving a specific operational requirement.

Currently permissible activities are shown in Step 1. In Step 2 it would be possible to see arrangements or legal gateways established for specific operational needs. Once established the CBD facility could be extended to support those operations (e.g. a government call-centre).

Longer term, the facility may support other demands such as those expressed by the HMT Banking Review (July 2000) which recommended that certain government information be released to the financial sector to improve the relationship between the financial sector and business. However, Step 3 would require substantial legal review and is likely to require new legislation to overcome issues of ultra-vires and to address privacy protection.

In practice the incremental approach - likely to be more substantive than three steps - would be a useful development strategy. It would provide the opportunity to test the constructs of the facility and ensure that the appropriate security and control mechanisms can be built in a safe environment.

The likely result of an incremental approach is that there will not be a single CBD but rather a series of CBDs developed to meet specific interdepartmental group needs. The prospect then exists for joining up the various CBD facilities as each incremental step is developed and brought into effect. The daisy chain of CBDs is represented in the diagram below:

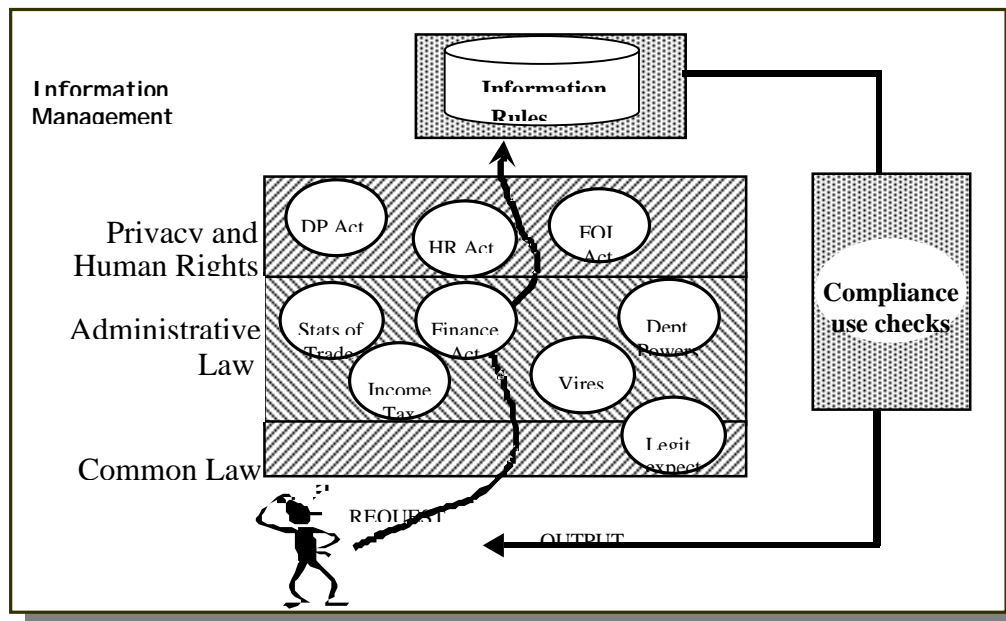


Introducing Change - The layers of law

The legal protection given to data held by government is quite complex and therefore not open to a single change that will suit all situations. This means that if legal changes are to happen they should be managed as a programme of changes that may meet short-term requirements but which are not independently determined. This incremental change informs data sharers and potential technology developers that any e-Government service will need to be built with the right level of flexibility to ensure that the correct level of data protection is given at each incremental stage.

The legal issues extend further than the duty of the Data Protection Act. To ascertain the extent of this, the project employed the services of a lawyer experienced in government centric legislation. In addition, to provide the broader understanding, views of protection and privacy professionals were sought. This included input from an e-democracy and human rights analyst.

The protection of people's rights can be traced through a number of legislative layers. At the simplest descriptive understanding there are four distinct levels to be addressed:



- Common Law - those rights developed by precedent whether applied to property or information
- Administrative Law (I) - regimes such as Income Tax, Company Registration which require specific actions of compliance and that determine how this will be conducted
- Administrative Law (II) - establishing specific roles of public bodies within which they shall act and not exceed these roles, act in ultra vires
- Privacy and Human Rights - a grouping that includes specific legal rights such as Freedom of Information, Data Protection, Human Rights and so forth.

The diagram below represents the complexity present, though the reader should note that the order of legal intervention is not quite so straightforward.

The diagram indicates that the problems lie deeper than the oft quoted Data Protection Act and begins with the legislation that determines the role and functions of a department. This is the first limitation to any form of data access and supply. If that path is negotiated then there are a number of privacy rules to be considered. The protection may be within the written Act - for example the Human Rights Act - or fall within common law precedence. The most significant example identified is that of 'Legitimate Expectation.'

Added to the above list is the policy structure which may be imposed to ensure that the specific operations are most effective (e.g. internal Codes of Practice) and one can understand that a blanket order to permit joined-up government is not a simple matter.

This is not a matter of 'civil service' bureaucracy that can be circumvented by a move to a private sector supplier. A relevant example relates to Canadian experience where two departments were brought together, and linked their databases. Though this seemed logical and expedient to improve services, public opinion against this form of data sharing won the day, and the databases were unlinked again.

It is difficult to see how one single legislative change could be safely proposed and implemented while positively engaging a whole nation.

Summary of Findings - Phase II Technical

The project concluded that it is feasible to build a number of virtual warehouses, and that this solution has many applications throughout government. It is a solution that is complementary to other work programmes for it supports the e-government frameworks, it is independent of organisational systems and is able to adapt to changing organisational and data structures.

The Technology Approach

The CBD model depends on two main technologies: virtual data warehouses, and probabilistic matching tools.

A **virtual data warehouse** provides transparent access to existing data sources, which are typically separate databases with different owners, and on different technical platforms. In essence, a virtual data warehouse enables integrated access to varied data sources, giving many of the benefits of a single centralised database, while retaining the flexibility of separate (and separately controlled) data repositories, and having a very low impact on existing operational systems.

Through a single connection, a virtual data warehouse can simultaneously connect to and use different client applications e.g. Oracle, Microsoft SQL Server, DB/2, Informix, Progress, CA-Ingres and other standards-compliant database engines. All the databases are treated as single logical unit, although they may exist on different computers in different departments, connected through the Government Secure Intranet (GSI).

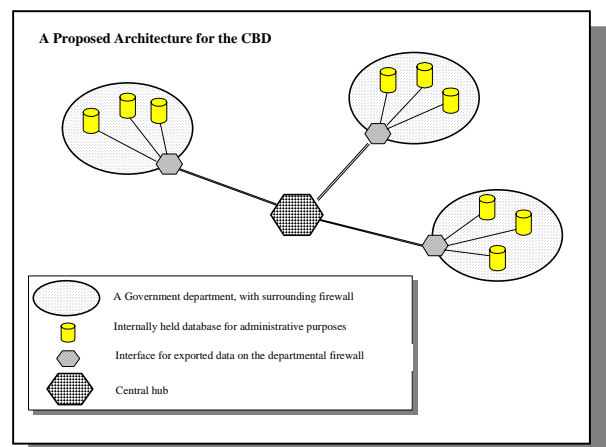
Internally, a virtual data warehouse has a layer of data consumers (to retrieve data from existing sources using a variety of protocols), and a layer of data producers (to provide data to client applications using whatever protocol is suitable). These layers are connected with an intelligent data integration layer that allows data joins between different sources. Virtual data warehouse products differ in the breadth of connectivity options, and in the richness and functionality of the data integration layer.

Probabilistic matching tools are needed, because Government data sources do not use the same identifier for businesses. The Single Business Register project evaluated the feasibility of a single identifier for all businesses for all purposes, and came to the conclusion that it was not achievable. Probabilistic matching tools enable linking of records in different databases based on fuzzy information such as company name, business address etc.

Proposed Architecture

The original intention was for a service that acts as a single central data repository. However, because of legislative structures, departmental targets and process-related complexity, the proposal is to establish the architecture for data unification through the use of intelligent hub technology. The proposal reflects on the findings that a Single Id for use in all circumstances is not achievable whilst promoting the rationalisation of numbering systems.

It is recognised that a single, monolithic approach would be inappropriate, and would stifle individual initiatives underway or planned by Departments. Through a unified programme a



common design and technical approach that develops best practice in data management processes can be achieved.

The proposal is not a single solution that can be situated in the centre of government but rather a solution that will be situated in the centre of the natural clusters that exist within the government (or public sector) universe. The first level of cluster may be a Department. The solution is a set of tools that can be used together to provide the opportunity to search, match, harmonise, de-duplicate and support data use activities. Pictorially it may be represented as below.

Meeting a Government Requirement

The delivery of a full knowledge base of economic activity in the UK is not possible from a single source or registration. This is not to minimise the importance of good registration systems - these are the lifeblood for many activities completed by public services.

Many businesses will fall into various clusters and will conform to a common type - either by size, purpose, or class. Within these clusters it should be possible to find a common reference or identifier that would suit the departments that will operate in or around each cluster. But across clusters there will be differences and these differences will lead to frustration through inflexibility or misrelated information.

The original goal of the project was to improve knowledge and to support joined up government. The legal and policy work has shown that it is difficult to provide such a service in a short time-scale. The technology solution is one based upon:

- ◆ managing the transition of legal change,
- ◆ managing disparate regimes of data handling and policy,
- ◆ integrating disparate data sources,
- ◆ supporting new e-government services,
- ◆ harmonising data and data usage within government,
- ◆ developing the benefit of private sector information sources to inform e-Government,
- ◆ providing departments with an internal data joining strategy,
- ◆ form a benchmark environment for migration to government e-Standards (e.g. e-GIF, e-GMS).

In essence it is a single focus that is rewarding in many areas for participating departments.

The rules must be very well defined and the intelligence used to determine data use would be of different complexity under different circumstances. For this reason, and not only due to the legal controls, a stepwise development within similar cluster environments is recommended.

Product Summary Assessment

The project evaluated probabilistic matching tools, and built a prototype demonstrator using a virtual data hub.

All the matching software reviewed is capable of providing facilities as required for a Comprehensive Business Directory.

The report cannot recommend particular products or particular suppliers for data hub or probabilistic matching tools, for this is a factor for each hub owner to manage through open procurement procedures.

Out-of-the-box performance in each case was high enough to commend further work. All the products showed the capability for further tuning. The products were also shown to be useful in harmonising and fault-diagnostic processes. They all offer very high degrees of flexibility. Due to the necessary trial limitations, it is not possible to state whether one package would be "better" than one of the others. The packages offer different ways of integrating with other systems, and different extra capabilities.

There are differences in principles, which may be more appropriate in different circumstances. This suggests that in a complex environment more than one product may be required. More detailed and intensive work would be needed within specific contexts to determine appropriateness of one over the other - or best combinations.

The intermediate recommendation is that any prototype system could use any of the reviewed matching packages; the determining factor as to which is used by other departmental participants is likely to be commercial availability, coupled with availability of appropriate skills and experience within a hosting department.

Summary of outcomes

- Matching data from different sources is possible. This can be done where there is a common key, but is also possible (with a lower "hit" rate) when only business name and address are held, using probabilistic matching tools.
- Matching data from different sources increases the value and richness of the data available.
- Tools exist to enable data to be linked dynamically across networks such as the GSI.
- Bringing in data from commercial sources such as Yellow pages adds a new dimension to the information available on businesses.
- The ability to see many views of the same attribute (e.g. industrial classification) across many data sources enables a more robust appraisal of the activity in which an enterprise is engaged.

Conclusions

The project team used professionals from the field of law and technology to determine the options for delivery - and these are applicable to data handling in general. Working in focus groups and technology partnerships the team concluded that:

- The complex UK legal structure presents a number of obstacles ranging from the administrative law through to privacy policy.
- Current legal gateways and commercial data sets provide an opportunity to make more effective use of data within the given limitations.
- A technology solution would need to be adaptive to changing legal powers
- A long-term vision of a single unifying service would need to begin with dispersed solutions that conform to a common model

- The technology exists to allow connectivity of disparate data sets without the need to replace or re-engineer legacy systems
- The technology available would assist services to be e-GIF compliant and improve downstream delivery.
- Combining commercial data with public sector sources provides a data and informational value.

Project recommendations

The project has concluded that it is feasible to build CBD virtual warehouses, and that the solution has many applications throughout government. The solution is complementary to other work programmes for it supports the e-government frameworks, it is independent of organisational systems and is able to adapt to changing organisational and data structures.

The legal and policy stage of the CBD project found that a widescale government-wide data sharing would not be possible without new primary legislation. This is an important finding that has implications for data sharing within government. It did however identify some areas where the CBD approach can be used beneficially.

The following recommendations are made:

- **Recommendation 1.** A Single Business Number supported by a Single Business Register is not pursued
- **Recommendation 2.** A technology configuration that can draw on disparate sources and match with accuracy can be developed and this should be adopted by departments to improve internal business data handling.
- **Recommendation 3.** A technology configuration that can draw on disparate sources and match with accuracy can be developed and this should be adopted where departments wish to and have the legal powers to relate data sets.
- **Recommendation 4.** A government information hub should be built using publicly available data (commercial or open public sector sources) and made available through GSI or related public sector infrastructure to improve knowledge and reduce the burden of independent departmental procurement and collection.
- **Recommendation 5.** It is recommended that a Government Information Management Strategy is developed to encapsulate the issues facing public sector closer working.
- **Recommendation 6.** A voluntary register could be established that, using OeE authentication services, could provide a reference source for a range of Government transactions and linked with a number of digital identifiers
- **Recommendation 7.** Legislative changes should be introduced but to avoid a public misunderstanding, the changes should be made incrementally to address specific operational (modernisation) requirements.