

Business Information In Government - Summary Report

1. Introduction	2
2. Project Conduct	2
3. Summary of Findings - Phase 1	3
3.1 Business referencing	3
3.2 Business registration	3
3.3 Storing and sharing of data	3
3.4 Rationalisation of business interactions	3
4. Summary of Findings - Phase 2: Legal and Policy	4
4.1 Currently permissible developments	4
4.2 Introducing Change - the layers of law	5
5. Summary of Findings - Phase 2: Technical	5
5.1 The Technology Approach	6
5.2 Proposed Architecture	6
5.3 Meeting a Government Requirement	7
6. Conclusions	7
7. Project recommendations	8

Business Information In Government - Summary Report

1. Introduction

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

It explored 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.

The following departments participated in the research with the Office for National Statistics providing project management and support: Companies House, Inland Revenue, HM Customs & Excise and Department of Trade and Industry.

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

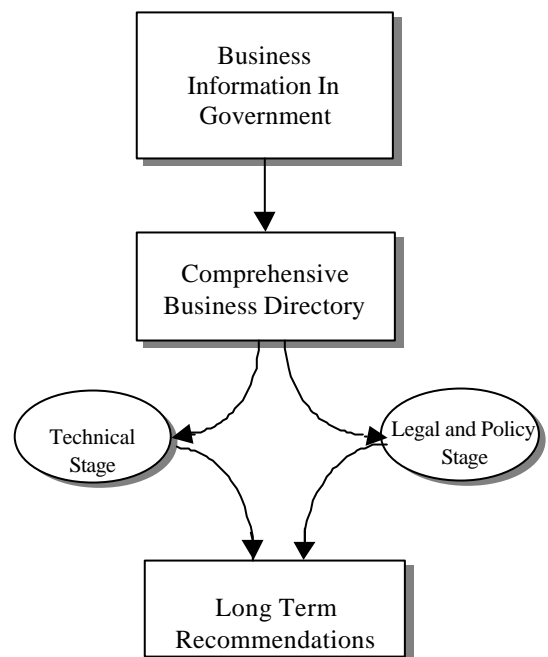
- Business Referencing
- Business Registration
- Business Data Management

2. Project Conduct

The project ran between April 1999 and August 2001 with the work split into two phases. At the end of the first phase (Business Consultation and Information Flow Modelling) the results were presented to the Information Age Champions and were published in the paper, *Electronic Service Delivery for Business*.

Phase 1 and the subsequent work concluded that it was not possible to create a unique number for all government purposes 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.



3. Summary of Findings - Phase 1

A number of significant conclusions have emerged from the Phase 1 work. Any consideration of business registration and referencing has wide implications for business and Government, and any business referencing scheme designed in isolation will be unlikely to succeed.

The main findings from Phase 1 are set out below:

3.1 Business referencing

A better means to identify and reference business across departments would be beneficial. The project concluded that 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. 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.

A better approach will be to use the existing referencing methods and to develop matching and recognition algorithms.

3.2 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, although attractive in principle, is unlikely to be practicable. In preference, there should be greater harmonisation between separate registration systems with a view to linking where this adds value.

3.3 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.

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.

3.4 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 common data. The immediate scope is limited by the difficulty of defining common data and confidentiality, and should focus on registration of information in some key business role(s).

4. Summary of Findings - Phase 2: Legal and Policy

4.1 Currently permissible developments

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

- Specific Administrative Purposes, Research and 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, but always operational, requirement.

Step 1 - Specific Administrative Purposes, Research and Subject Access

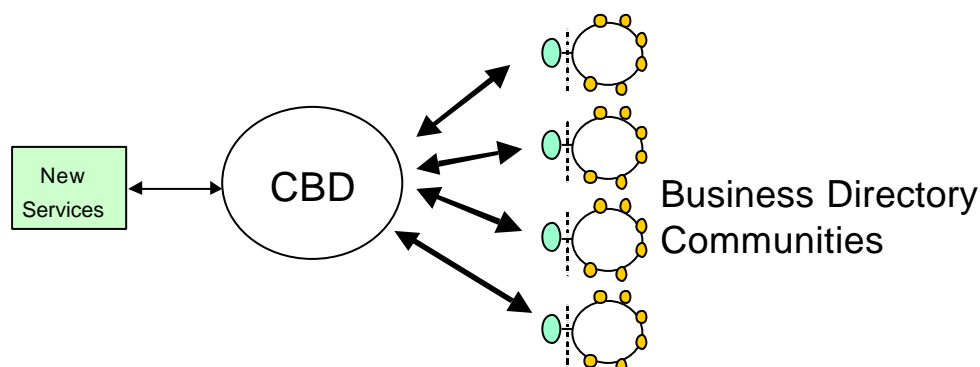
Step 2 - Call Centres and Joined-Up Processes

Step 3 - Special Arrangements and Public Listing

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. 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 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 business directory (BD), but rather a series of business directories developed to meet specific inter-departmental group needs, linked to the core CBD. The prospect then exists for joining up the various business directory facilities as each incremental step is developed and brought into effect. The linked business directories are shown in the diagram below:



4.2 Introducing Change - the layers of law

The legal protection given to data held by government is quite complex and therefore not currently 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 sought independent legal advice from experts in government 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 that require specific actions of compliance and 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.
- Privacy and Human Rights - a grouping that includes specific legal rights such as Freedom of Information, Data Protection, Human Rights and so forth.

The problems lie deeper than the Data Protection Act and begin 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. There are also 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. The projects view is that currently a single legislative change could not be safely proposed and implemented.

5. Summary of Findings - Phase 2: 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 complements other work programmes, as it supports the e-government frameworks, is independent of organisational systems and is able to adapt to changing organisational and data structures.

5.1 The Technology Approach

The CBD model is not a directory as such, but more a mechanism for linking together datasets to provide a joined-up view of information. This 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, whilst retaining the flexibility of separate (and separately controlled) data repositories, and having a very low impact on existing operational systems.

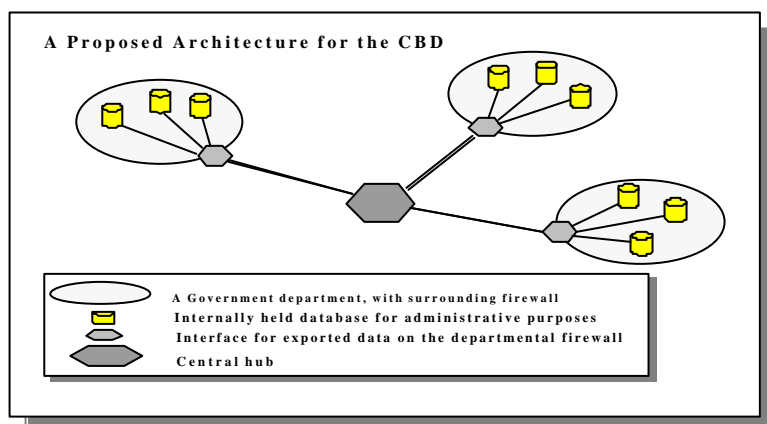
Through a single connection, a virtual data warehouse can simultaneously connect to and use different client applications, e.g. Oracle, Microsoft SQL Server, DB/2, Informix, Progress, CA-Ingres and other standard database engines. All the databases are treated as single logical units, although they may 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), and a layer of data producers (to provide data to client applications). These layers are connected with an intelligent data integration layer that allows data joins between different sources and controls access within a prescribed system of regulation.

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. These linked records increase the value and richness of the data available, and the ability to see many views of the same attribute (e.g. industrial classification) across many data sources enables a more accurate appraisal of the activity of an enterprise. Additionally, bringing in data from commercial sources such as Yellow Pages adds a new dimension to the information available on businesses.

5.2 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 has been to establish data unification through the use of intelligent hub technology. A virtual data hub has been developed that brings together data from the administrative sources and



Yellow Pages, and provides a cross-government view of information. This proposal is based on the assumption that a single identifier for use in all circumstances is not achievable.

It is recognised that a single, monolithic approach would be inappropriate, and would stifle individual initiatives underway or planned by departments. 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 government (or public sector). 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 and support data use activities.

5.3 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.

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.

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 systems of data handling and policy and integrating disparate data sources
- supporting new e-government services and forming a benchmark environment for migration to government e-Standards (e.g. e-GIF, e-GMS)
- harmonising data and data usage within government
- providing departments with an internal data joining strategy

6. Conclusions

The project team used expert advice on legal and technology issues 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 provides a number of safeguards ranging from the administrative law through to privacy policy.
- Current legal gateways and commercial datasets provide an opportunity to make more effective use of data within these safeguards.
- A technology solution needs 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.
- A voluntary register could be established that, using Office of the e-Envoy authentication services, could provide a reference source for a range of government transactions, linked with a number of digital identifiers.
- Augmenting commercial information with public sector sources provides added value to the data.

7. Project recommendations

The project has concluded that it is feasible to build a CBD virtual warehouse(s), as has been done in the ONS, and that the solution has many applications throughout government. The solution is complementary to other work programmes for it supports the e-government frameworks, 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 extensive government-wide data sharing would not be possible without new primary legislation. It identified some areas where the CBD approach can be used beneficially.

The study has resulted in the following recommendations :

- A Single Business Number supported by a Single Business Register is not pursued.
- A technology configuration that draws on disparate sources and matches with accuracy should be developed and adopted by departments to improve internal business data handling.
- A technology configuration that draws on disparate sources and matches with accuracy should be developed and adopted where departments wish to and have the legal powers to relate data sets.
- 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.
- It is recommended that a government information management strategy is developed to encapsulate the issues facing public sector closer working.
- Any voluntary registration service developed by the Office of the e-Envoy should be developed in line with the principles of the CBD project.
- Where legislative changes are made affecting issues of data access, then the approach should be developed in line with the principles of the CBD project.

The first four recommendations have been taken forward within stage 2 of the Comprehensive Business Directory project, which has initially been running from November

2001 to September 2002. The remaining three will be addressed by the CBD steering group. This group, which comprises members from Companies House, Inland Revenue, HM Customs & Excise, DTI Small Business Service and the Office for National Statistics is managing the ongoing programme of the CBD work.