Case Study: Business Data Management

The Application-focused organisation

Traditionally organisations have seen themselves as being application-focused. The organisation took in data from a variety of sources – sensors, manual data collection, buying in external data etc, and used that data to help provide services (and or products) to its customers, or staff. Raw data was obtained from its sources by applications (either built in-house or purchased) and applications delivered information. In general the applications receiving the data and those outputting the information would be different requiring some form of data transfer between them.

Many of the applications would be commissioned by individual departments (or teams) within the organisation. These commissions would therefore be focused on the delivery of some specific service and would typically treat any internal data management and transfer as technical detail of secondary importance. As a consequence data would be trapped within vertical application silos and unavailable for widespread utilisation across the organisation, as illustrated below. In addition bespoke development created non-standard data transfers between applications locking the organisation into long-term dependency upon the application vendors.

This siloing effect also resulted in replication of effort as the same data might be collected by multiple applications and similar information delivered to multiple users.

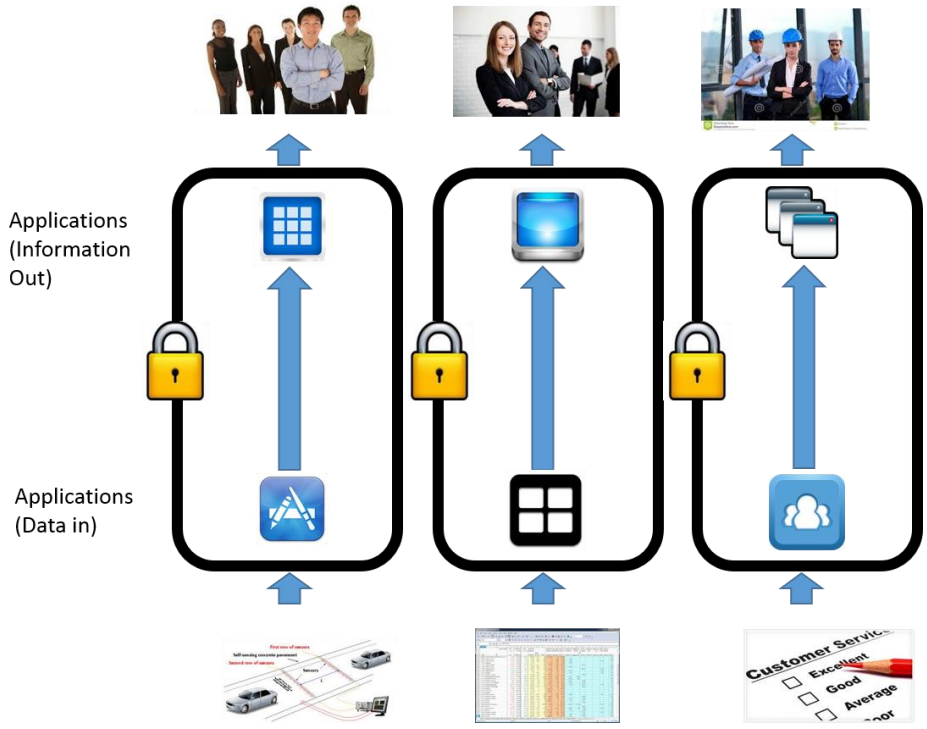

Figure 1: Application silos

The Data-focused organisation

Increasingly, organisations are beginning to realise that the data they collect and curate is one of their most valuable assets. (Major companies such as Google, Amazon, Facebook, Apple and others realised this some time ago and give away free services in an effort to collect and become the managers of as much consumer and business data as possible). Organisations who maintain effective control of their

data are able to find new innovative ways to exploit their data equity and at the same time to free themselves from vendor lock-in.

Ensuring that data collected by applications is maintained with the organisation and that applications delivering information to customers and/or staff obtain their data from the organisation (rather than directly from another application) breaks down application silos increasing the potential exploitation of data and making it easier for the organisation to replace applications if required.

This is illustrated below. Data collected is held within a horizontal Data Management Layer and then made available to applications as required. The availability of a broad range of data enables new and exciting applications to be developed which can extract value from disparate, even seemingly unrelated, data sets. The decoupling of applications makes the entire system easier to understand and enables one application to be more easily replaced by another.

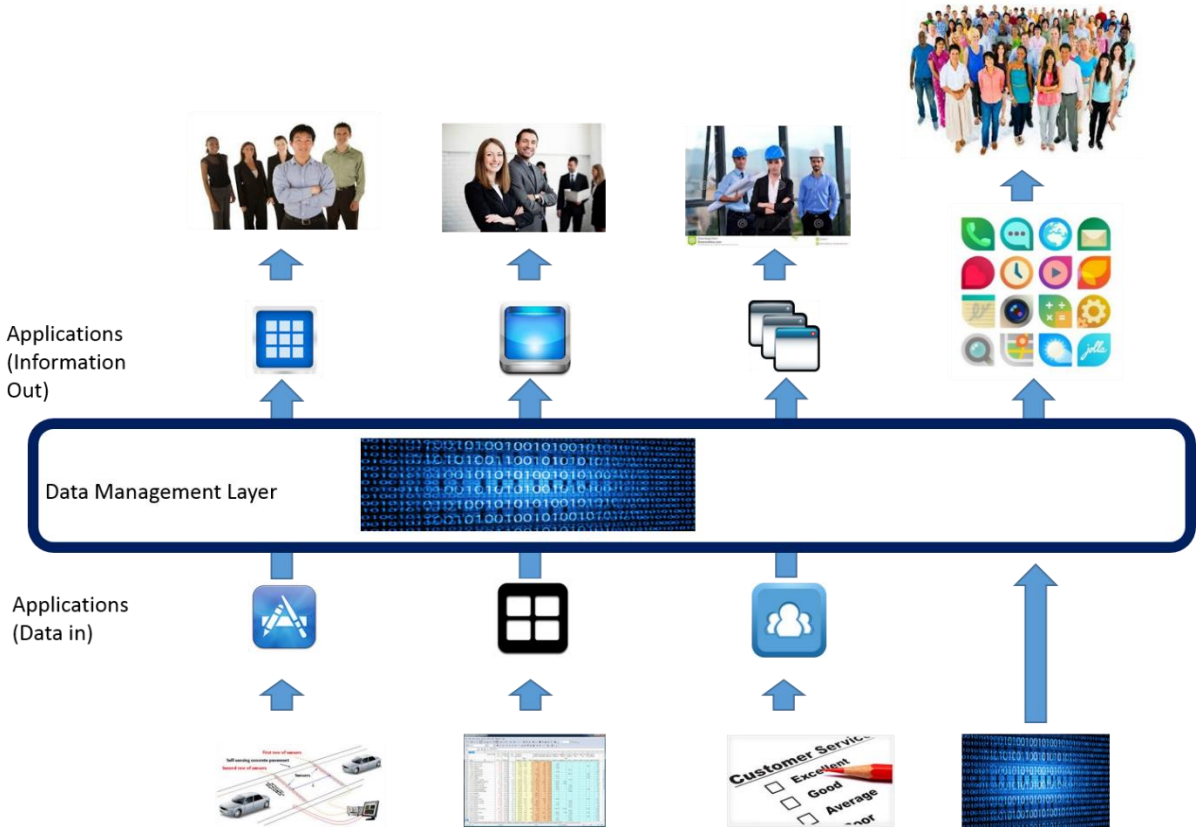

Figure 2: Freeing up data

The case study which follows describes an example in which a Data Management Layer has been introduced to Newcastle University. The example illustrates that adding a Data Management Layer need not be a disruptive process. It can be achieved incrementally and can produce immediate benefits which encourage further progress.

Business Data Management

Many organisations obtain similar or related data from multiple sources. This can result in different data sets, purporting to hold the same information, being utilised in independent silos. Where the information held in the data sets differs this can lead to inconsistencies and inefficiencies within the organisation. The fact that data is locked within silos frequently prevents the organisation from correcting, or even detecting, these problems when they arise.

As one example, Newcastle University, in common with many organisations, collects information on the organisations it engages with (or would wish to engage with) from many different sources. In the case of Newcastle University sources include data provided by the University's Business Engagement teams, data collected from surveys and events, data purchased from Companies House and from commercial Business Data providers. All of these sources can contain invalid information. Even Companies House data, often considered the 'gold' standard, can be out-of-date or incorrect due to data entry problems.

The use of the different sources by different applications (utilised by different departments) can be problematic. For example: a Business Engagement team might deal effectively with an organisation but the Careers team might hold an invalid address for the organisation and therefore fail to effectively coordinate their engagement; one department might engage with 'J. Jones and Sons' and another with 'John Jones Ltd' without realising these are the same business.

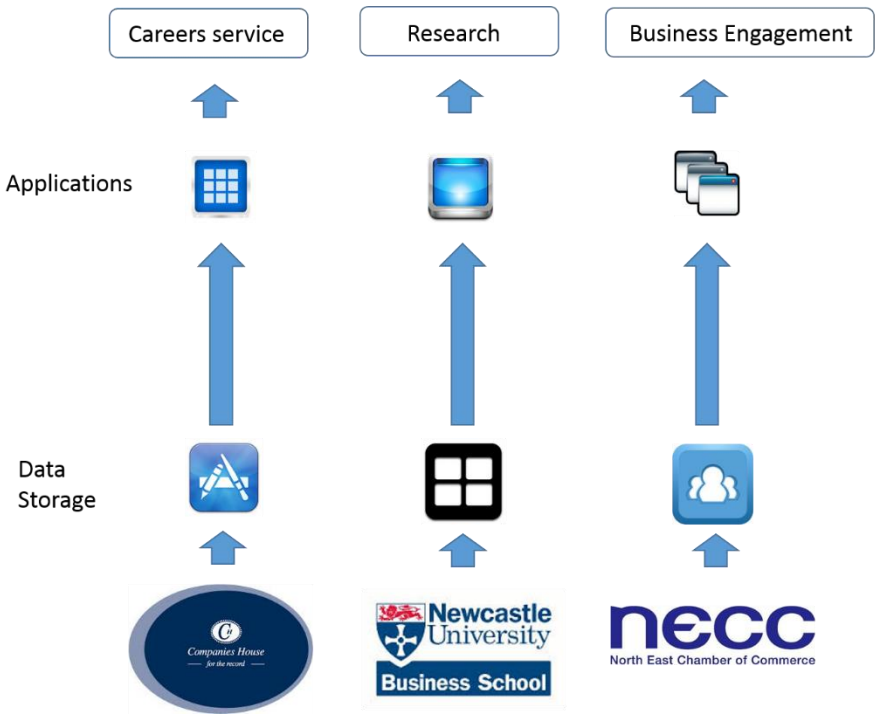

Figure 3: At the start of the project

Identifying these forms of inconsistency is a notoriously difficult problem, made well-nigh impossible by the existence of siloed information as data is collected independently, and is held and output in different formats. Even where inconsistencies are spotted, often following a problem having arisen, correcting the data is a manual error-prone process which needs repeating again and again each time new data is ingested.

Newcastle University turned to Arjuna Technologies for help.

The Agility DataBroker

Arjuna Technologies have created their Agility DataBroker product to act as a Data Management layer in just such a scenario. The project wanted to proceed carefully with minimum disruption to the various teams so as a first step introduced the Agility DataBroker as a means of holding copies of the data collected by each independent system. To do so they created a plug-in for the Agility DataBroker which was capable of obtaining the data from the existing Data Storage systems. They then created a set of Agility DataBroker plug-ins capable of analysing the data and outputting a 'Data Inconsistency Report' identifying any inconsistencies discovered. This report could be used by University teams to assist them in correcting their data sources, or in making allowances for the inconsistencies within their internal processes. Note that the report (or a version of it) could be delivered to the organisation from which original data was obtained allowing them to clean the data themselves. For example during this project it became apparent that there were significant errors in the NECC data (generally created by transcribing data from paper forms). The report generation was implemented as an automated process so that the ingestion of new versions of any of the data sources could generate a new report. An interim architecture after the introduction of the Agility DataBroker is illustrated below.

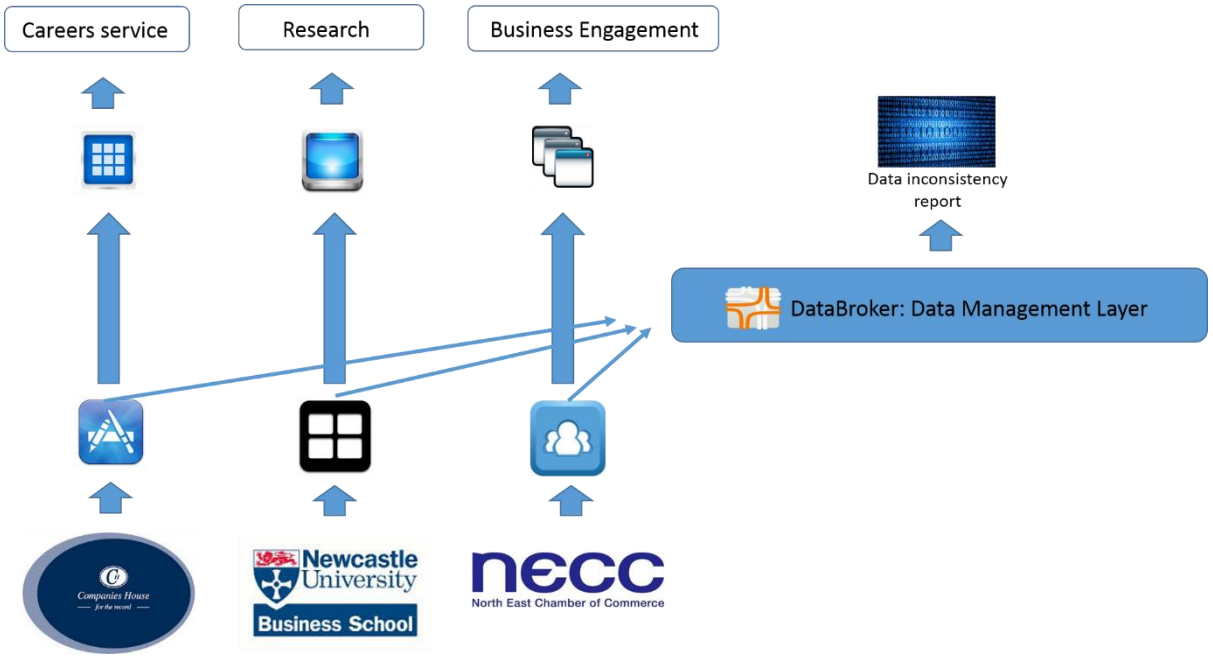

Figure 4: An interim solution

The introduction of the Agility DataBroker and the publication of the 'Data Inconsistency Report' was a relatively simple process which required minimal effort and no disruption of the existing systems. However, this first step did little to address the fundamental inefficiencies caused by the existence of multiple application silos. Addressing these issues requires a further series of incremental steps by which the Agility DataBroker would be extended so as to deliver the data collected directly to each of the applications delivering information to users. Once this has been achieved the means of collecting and storing the data can be rationalised so that duplication and inconsistencies can be removed.

This solution also allows the data to be made available to other applications and therefore encourages the creation of new innovative services as well as improving the efficiency of existing ones.

The eventual, intended architecture is illustrated below.

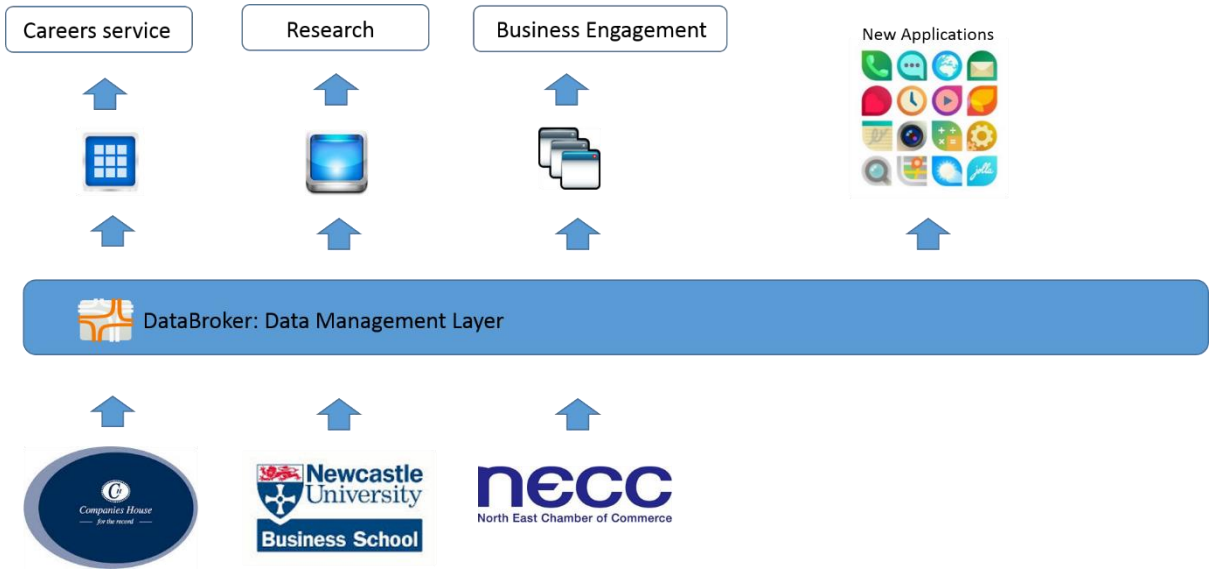

Figure 5: The eventual outcome

The project illustrates the problems with allowing data to be consigned to application silos and the advantages of introducing a Data Management Layer. In this case the introduction of the Agility DataBroker enables the inefficiencies in data collection and storage to be removed, and the ability to deliver the data in multiple outputs opened up new opportunities to share the data within the organisation and beyond. Sharing the data and combining it with other data allows new information to be obtained with the potential to improve decision making thereby reducing costs and/or improving service.

The Agility DataBroker was introduced without impacting upon other applications and it should be clear that a step by step expansion in order to consume additional data would continue to provide additional benefits. In fact with data the sum of the total exceeds the sum of the parts as adding more data sets increases exponentially the number of ways in which data can be combined. In addition the Agility DataBroker is capable of delivering data in whatever format is required and with whatever transformation is required. This opens up the opportunity to deliver data to both internal and external application developers in order to encourage innovation, and to citizens and improve access to Council information.