

Human and Technological Resources Needed to Develop and Sustain A City-Wide Educational Data Observatory

Rachel Shanks
University of Aberdeen
orcid.org/0000-0002-7895-5136

Bruce Scharlau
University of Aberdeen
orcid.org/0000-0003-4646-5973

Hataichanok Saevanee
University of Aberdeen
orcid.org/0000-0003-1630-2380

Kevin Stelfox
University of Aberdeen
orcid.org/0000-0002-4353-5262

Abstract

A data observatory for a local government authority in Scotland was developed in collaboration with university researchers and other stakeholders. The data observatory provided a platform to extract, analyse and view data from various databases. The data observatory was created using standard web tools. One component of the observatory, the DataHub, is provided as a case study. It was developed in partnership with other agencies and produced new information on school leaver destinations. After two years the project paid for itself in terms of staff time saved. Lessons learned about setting up a city-wide data observatory are provided.

Introduction

This chapter focuses on supporting data analytics in education through human and technological resources. A data observatory for a local government authority in the North East of Scotland was developed using a research grant under

the supervision of researchers from the University of Aberdeen. A developer worked with local government authority analysts to develop a system software platform using standard web tools connected to a local government authority education attainment database. This meant the database could be adapted and expanded as the team learnt more about what was needed, and what was possible. Information about each school in the area was collected in a standardised way. After two years the project paid for itself, in terms of staff time saved, and the local government authority extended the developer's post indefinitely so that the system platform could be developed further.

First, background information on the data observatory is provided, then resource support issues of data protection, data validation, data access and user-authentication are explored, including the use of data by non-specialists.

Technological aspects of the project including automation, in-house programme development, and multi-partnership working are also covered. Where appropriate, the different perspectives of the stakeholders involved such as IT specialists, educational policy makers, school administrators, and researchers are provided. Finally, details of a subsidiary project, the DataHub is used as a case study. This project aimed to support those working with young people to help them into education, training, or employment.

Data Observatory Concept

A data observatory can be understood as a technological telescope to view stars (data) in greater detail or in constellations as it is up to the viewer to decide where to point the telescope at and what level of magnification to use. A data observatory is also a type of information system which can be defined as:

“A system to support the administration of environment, resource and planning tasks by making information available to executive and public through co-ordinating existing systems and investments by a common architecture” (Gunther, 1998).

A data observatory is an online resource that provides easy access to a rich source of information for different types of users. The data observatory concept has been applied in various types of organizations, including the National Virtual Astronomical Observatory in the USA, funded by the National Science Foundation and NASA (Brunner, Djorgovski & Szalay, 2001). The National Virtual Astronomical Observatory is an online research facility, which provides space and ground-based information to professional astronomers and the public. There are at least 19 astronomy observatory projects worldwide in countries such as Argentina, China, Germany and Japan (Djorgovski & Williams, 2005). In the UK there are data observatories such as the UK Soil Observatory (NERC, 2017), while in Scotland there is the Scottish Public Health Observatory (2014), which provides evidence and statistics on public health. Across the UK there are several local government authorities that have data observatories to help in the performance of their functions. There are also UK regional data observatories bringing together the data from local government authority data observatories such as the regional observatory for South Yorkshire (Fluent Technology, 2012). Using a Local Information System, South Yorkshire Regional Partnership, reported an annual saving of approximately \$100,000 from a reduction in staff time managing enquiries from the public and a 50% reduction in time required to produce national statistical reports (Fluent Technology, 2012). Several data observatories not only incorporate local

government authority data, but also include health, and police service data in order to provide a more detailed regional picture (Foley, Alfonso & Wiseman, 2007).

The data observatory explored here provided each of the local government authority's primary and secondary schools with a profile page. This covered the age, ethnicity, absence rate, attendance, attainment levels, and predicted grades of all pupils. The profile page also included the corresponding data for pupils' home addresses from the Scottish Index of Multiple Deprivation (SIMD) which provides data on the levels of socio-economic deprivation by geographic area (Scottish Government, 2017, November 03). This information could be compared with the other local government authority schools in the city via graphs, charts, and tables. This meant school staff were able to look at school data for that year, and over time, to support evidence-based decision-making. School management teams were able to use the observatory data to provide contextual, background information to the national school inspectors, Her Majesty's Inspectorate of Education (HMIe).

The aim of the project was to make better use of the vast amounts of data held by the local government authority in disparate, and disconnected databases, and data sets. Much of the data were hard to find, and access, even if it was available in principle. The local government authority received a high number of requests for information in a form it was not able to provide. By enabling the combination, and unlocking of data, and by making it freely available to decision-makers, and service-users, comprehensive, robust and evidence-based decisions could be made on performance, policy and planning. By having consistent data available to all partners involved in school education in the city, decision-making was enhanced, and enabled cost savings through the reduction of duplicate data collection, maintenance, and storage. The data observatory added value to the

information already gathered by providing data interpretation, presenting aggregated data at a geographical level or collating data. The data observatory also provided an opportunity to support collaboration through partnerships. In addition, in future, it could enable the public to access information themselves, thus reducing the cost of dealing with individual requests which is both time consuming and resource intensive.

The local government authority, in partnership with researchers from the University of Aberdeen, recognised that effective performance monitoring, planning, and policy decision-making, required a robust evidence base from which to make decisions for monitoring existing services as well as developing future service developments and innovations. The local government authority and partner organisations understood they had all the necessary information, however, it was not always readily accessible. The data observatory provided the basis for a solution to join data from a range of sources and make them accessible to the appropriate personnel in a suitable and convenient format. The platform removed the drudgery of data manipulation, thus freeing up staff time for higher level analysis work. The data observatory enabled data to be more accessible to local government authority department officers, partner organisations, and in time, members of the public. Information was provided that was up-to-date and 'live' with the provision of accurate and consistent data and reports, including trends in key indicators, for decision-makers.

Conceptual model

The data observatory provided a robust, secure platform to store and collaborate on data with requisite data analysis and correlation tools. There was a

growing interest in the data observatory usage as more staff became aware of what was possible with the platform. Staff were interested in how data sets had been combined, and its ability to overlay data sets onto maps to create heat maps showing areas of greater concern. This raised the interest of the national skills body, Skills Development Scotland, to see how they could use this for their data sets across all of Scotland.

The data observatory conceptual model in Figure 1 illustrates how the components fitted together. Data were retrieved from the Management Information System (MIS) database (DB) and pulled into the data observatory where it was gathered and fused with other data sets before staff could query it, as needed, for their work. The users were local government authority staff, and in future it was hoped to be extend this to partners, researchers, and the public.

[Insert Figure 1 Data Observatory conceptual model here]

The data observatory delivered the functionality to load, store, analyse, and present statistical data. It provided the ability to present this data using visualisation tools like charts and graphs, and the facility to export it into PDF format. Information could be in various formats such as XML, Excel, and CSV. Dynamic reports were created with interactive maps, charts and tables together with time series. Key indicators could be shown based on user requirements. These features provided the ability to produce strategic assessments based on a combination of maps, graphs and contextual analyses. The data observatory provided the opportunity of transforming the way the local government authority processed and used data.

In the next sections, resource support and technological support issues are discussed and the DataHub case study illustrates the potential of the data observatory.

Resource support issues

The development of the data observatory provided a methodology for developing software for the collection and presentation of data from a wide range of sources. With this methodology, it was possible to use the approach on as wide a range of topics as was necessary. The deployment of the developer within the team of analysts ensured that the development of the data observatory was focused on the needs of the staff in relation to their data requirements. Development specifications were drawn up by the developer-who organised, and facilitated a number of workshops in collaboration with staff, for the personnel who accessed and used that data. These were later followed up with one-on-one discussions with individual staff to elicit specific details. The typical users included the existing research and intelligence teams, the policy and performance officers, elected representatives and leaders, and school management teams. At the time of writing, data observatory access for the general public is still under development.

Data protection

Legal requirements relevant to the data used in the data observatory, such as the Data Protection Act 1995 in the United Kingdom, address confidentiality and privacy issues. The team faced challenges working within the legislation, and so collaborated with the authority's legal team. Data protection principles from the legislation specify that information must be fairly and lawfully processed; processed

for limited purposes; be adequate, relevant and not excessive; be accurate and up-to-date; not kept longer than necessary; be processed in line with the rights of individuals; be kept secure; and not transferred to other countries without adequate protection (Data Protection Act, 1995). To ensure data security, the data were stored centrally, and a data sharing protocol was established.

Data authentication

Authentication was improved through close working relationships, and an incremental and iterative process between the developer and data analysts ensured that formulas were correct and errors were spotted. Thus, the data observatory helped to maintain the quality assurance of the original source data. No assumptions were made regarding the accuracy of the data. The data was run in search of anomalies (e.g. 50% absence for a school on a particular date), to identify a problem with a formula, or the inputting of the original source data. On several occasions errors were found in the original data.

It was important to ensure consistency in the format and structure of the data otherwise a lot of time could have been spent in cleaning, and reformatting the data into a useable form. By giving users a unique password, the development officer was able to monitor and audit their activity, and work with them to improve data accuracy.

Data validation

Every year the Scottish Government releases school data by local government authority as a spreadsheet with columns and categories. These categories change on a regular, but ad hoc basis, which means that local government authorities must translate the categories to ensure the numbers in their

data sets make sense, and tally correctly. It is necessary to extrapolate what is there, and to understand what is missing. A rigorous process is necessary as formulas, and data are not always present. Schools provide data on pupil registration, and attendance via the SEEMiS software (SEEMiS, undated), and this is sent directly to the Scottish Government. Then, the Scottish Government sends the data to the local government authority, that must then authenticate, validate, and organise the data.

When updates for the school pupil data arrive at the local government authority, they are usually in the form of a Microsoft Access database or an Excel spreadsheet. The data were then checked to see if the columns had changed, and if other modifications were needed. For example, one year a new column appeared as 'COST_centre_key'. The developer found that the column related to the previously used 'school_code', which was missing, so she converted the 'COST_centre_key' to 'school_code'. When the extrapolation and modification was complete, the data were uploaded to the data observatory. From there, the algorithms were confirmed and approved by an analyst specialist and the new data set was made available to stakeholders.

Data access and user authentication

Approximately 150 individuals had full access rights to the underlying data in the data observatory. Others had access to the summary data that did not contain individuals' information, subject to relevant permissions and data sharing protocols. Due to the sensitive nature of the data in the data observatory, protocols were developed to determine who could have access to the data, as it could be drilled down to individual pupils and other young people. Access was limited to relevant local government authority staff, and school principals. This was in line with the

standard used by the Scottish Government Statistics Group which is that it must not be possible to identify individuals from datasets and/or analysis (Scottish Government, 2017, November 28).

People, both inside and outside the local government authority, had to have authorisation and clearance through their organisation to have access to the data. Those with full access to the data observatory could see the granular data at the individual level, for example that a particular pupil was absent from school on a particular day. General material, namely aggregated data, was available to anyone working in the local government authority, including numerical data and graphs. In the future, public access to the aggregated, anonymised information may be granted. By allowing the public access to this data, it should reduce Freedom of Information requests, thus saving staff time and resources. Access by the public highlights the issue of access to, and use of, the data by non-specialists. Furthermore, public access would enable access for researchers. This could lead to innovative policy initiatives based on the combination of reliable and consistent data, as well as new visualisations affording new insights.

It was important to have different levels of security clearance depending on the sensitivity of the data. It was possible to trace and store information on users. For example, we were able to identify what data were interrogated, by whom and how often it was accessed to discover who was looking for what data. This information helped to continuously improve the platform of the data observatory.

Use of data by non-specialists

Potential access by the public highlighted the issue of access and use of the data by non-specialists. There are challenges in allowing non-specialists, and other

people who do not understand how to interpret the data, and who, thus, might misuse the data by wrongly repurposing it (Nielsen et al., 2017). IT specialists need to make the software as easy to use as possible, ensure its correct use, and enable the visualisation of the data and fungible use of the materials to reduce this issue.

While the data observatory sat within the local government authority's Education and Children's Services Department, it was also an IT project. Consequently, the project was provisioned by IT but was maintained by the developer in the Education and Children's Services Department. This unusual arrangement, though sometimes challenging, enabled the unit to maintain freedom of operation with updates to the data observatory determined by the department, and not bound by the availability of IT staff.

As user needs changed, it was important to continually update the data observatory. If the data were presented in such a way that users could not understand or utilise it, the observatory would not be fulfilling its role. Making the observatory more accessible and user-friendly allowed a wider range of people to use it effectively. Giving users access to the data in different ways enabled people with different skills, and abilities to access the information, and provided flexibility in terms of data presentation. As well as providing guidance on the system, users were trained in how to make presentations using the data observatory data, and feedback from this activity was positive.

The next section addresses the technological aspects of the project. It covers automation, in-house software development and multi-partnership working.

Technological resources

The data observatory was developed using web-based applications. This arrangement had a number of advantages:

- No installation was required, thus saving time and resources in relation to setting up, support and maintenance.
- All types of users could access the observatory at anytime from anywhere using computers, smart phones or tablets.
- The web application could work on different internet browsers and was compatible with most operating systems.
- It saved space on computers as the software was run from the server and updates only needed to be done on the server thus users did not need to worry about whether the application was up-to-date.

Despite the benefits, there were also disadvantages associated with using a web-based application:

- The internet was required to access the data observatory
- The application could take time to run if internet connectivity was slow
- More care was needed in relation to security

Automation

It was not possible to automate all the processes first envisaged. However, some automation was possible such as the validation of data and uploading data to the platform. As a result, the estimated savings in terms of staff time for analysis and presentation of data were approximately \$175,000 per year. This included a reduction in staff time spent on school roll forecasting, wider achievement data collection, school profile data analysis, and integration and visualisation of previous

data sets. These savings more than paid for the staff costs related to the developer position.

Automatic updates and notifications by email to the school or other data provider to update information on a monthly, quarterly or annual basis ensured the information was up to date and useful. Previously this data was static and generally only accessed annually via paper reports or Excel spreadsheets.

After quality assurance staff were introduced to the data observatory, they asked for a report generation tool. This request highlighted how the data observatory was continually being developed to aid more people in their work. An example of how the quality assurance staff utilised the data observatory is as follows: they would select a school that they were working with and by selecting the school all the school data appeared as a web page. The data included the most current information such as pupil achievement levels broken down by 'free school meals' (a rough measure of deprivation), 'looked after children' (those in state care or being monitored) and levels of deprivation statistics by geographic area. The webpage could then be downloaded as a report. This report was a Word document thus enabling staff to edit the report.

In-house programme development

Originally, the research associate developer was going to produce a prototype data observatory and the local government authority IT specialist staff were going to develop and manage it themselves but in the event IT staff did not become involved in the project. IT staff only helped by allocating space on the intranet server along with the requisite web servers, and databases to which the data observatory could be deployed.

IT issues were overcome by the team working on and deploying their own solutions to their challenges, and by finding the simplest way to illustrate an issue to gain co-operation from others. The developer found open source and other freely available building blocks with which to build a solution. This meant using MySQL for the database, alongside the Microsoft IIS web server, and the community edition of Visual Studio to develop the data observatory web application. Ensuring the website was user-friendly was important, in terms of readability and usability to help the visually impaired.

Back end users were split between the data analysts of the research staff, information officers, and the IT team. The research and information officers assumed responsibility for what data were included and at what geographical level, its design (completed in conjunction with graphic designers) and for uploading it, whilst the IT team were responsible for the software hosting.

Multi-partnership working

Teachers, school administrators, local government authority departments, and organisations such as the national skills body, collaborated on the data observatory. For the data observatory to become a 'platform' for others outside the local government authority, IT specialists would have to develop a way for individuals from outside organisations and project partners to have access to what is a closed intranet.

The data observatory team collaborated with university partners to provide a space for interns to gain experience as programmers while working under the supervision of the developer. This enabled ideas to be prototyped more easily than would otherwise be the case. Over two summers a group of students worked as a

team to enhance their skills while developing a larger prototype of the data observatory. Each iteration found its way into the application at a later stage.

The data sharing protocol had different levels for data providers:

- Level one – we commit to share
- Level two – we commit to share this data
- Level three – we commit to share this data in this way

This multi-partnership working was exemplified in one particular case study which also highlighted the usefulness of the data observatory, namely the DataHub.

DataHub

The DataHub was a sub-project of the Data Observatory that was set up to gather and analyse data on young people who might need support as they were not in education, training or employment after completing secondary school. The DataHub data could be plotted via heat maps to identify clusters, and to enable local government authority staff to organise events where they could talk to people in that category and find out how to help them with targeted support. For example, to organise tailored career information events in specific geographical areas.

The data observatory enabled the local government authority to have more knowledge on the destinations of young people after they left school. School profile data were mapped to assist stakeholders in understanding school performance from a wide range of sources. Research and data teams across partner organisations could create much more detailed and frequent analyses, such as seasonal trend tracking, to explore the impact of changes to the local economy on the availability and choice of post-school destinations.

Data from Skills Development Scotland (national skills body) were incorporated into the data observatory to improve the support provided to young people to enable them to move into positive destinations after school (Skills Development Scotland, 2017). Data were updated on the first day of every month via the data observatory, and a snapshot of the available data on that day was produced. Further snapshots could be provided on request subject to agreement from other partners.

[Insert Figure 2. DataHub Case Study Model here]

At the strategic planning level, the information gathered through the interactive tools provided in the data observatory DataHub subproject was intended to assist local government staff and partners in their planning and assessing of service provision and needs in key geographical and themed areas. While at an operational level, data could be used to provide staff with additional details about the individuals they were (or were about to be) working with.

This model could be used by local government authorities across Scotland if standalone software was developed or each local government authority made use of webtools in the same way as the Data Observatory.

Conclusion and Next Steps

The data observatory project is a working sample of what can be done with data to enable transformational gains in both digital delivery of services, and information. The driver behind the data observatory was automation in order to reduce costs for the local government authority in a period of public funding cuts. However, other benefits were realised, such as improved quality assurance of the

data, as well as an increased ability to analyse, and compare data between schools, and against local and national benchmarks, by joining data that was not linked before, to analyse more data sets in greater depth. The team learnt that there was always more data that could be added to the platform, and that new data brings new understanding and knowledge to all users.

Several key lessons were learnt while developing the data observatory platform. Firstly, it was important to gain insights from stakeholders about what was needed early on in the process. Secondly, it was useful to provide a minimal version as soon as possible in order to gain credibility, and to let people see what was possible. Thirdly, it was important to inform as many people as possible about the data observatory and how they could make use of it. Finally, it was very useful to embed the developer with the analyst team so that they could learn from each other as they worked together and be in constant communication.

Summary

The support and technological resources required to develop an education data observatory have been laid out. The resource support issues included the protection, authentication and validation of data, alongside data access, user authentication and the use of data by non-specialists. The technological resources related to automation, in-house programme development and multi-partnership working. A sub-project of the data observatory, the DataHub, was explored to show the possibilities afforded by the development of an education data observatory. Recommendations are provided on setting up such an observatory.

Questions to discuss

1. If you were setting up an educational data observatory for your community or organization, how would you gain buy-in for it?
2. What is the most exciting domain with which you would start your data observatory? (And which could become a showcase to get others excited?)
3. What kind of data would you want to gather for your educational data observatory?
4. What type of activities would you use to promote your data observatory to develop support from your potential users?
5. What would you include in a data-sharing protocol with partner organisations?

References

Brunner, R. J., Djorgovski, S. G., & Szalay, A. S. (2001) Toward a National Virtual Observatory: Science Goals, Technical Challenges, and Implementation Plan. Retrieved from *arXiv preprint astro-ph/0108115*.

Data Protection Act 1995. Retrieved from http://www.parliament.scot/Fol/Data_Protection_Policy.pdf

Djorgovski, S. G., & Williams, R. (2005). Virtual observatory: From concept to implementation. Retrieved from *arXiv preprint astro-ph/0504006*.

Fluent Technology. (2012, June 27). LASOS South Yorkshire Regional Partnership. Origin Local Information Systems. Retrieved from http://www.localinformationsystems.co.uk/_Documents/LASOS.pdf

Foley, P., Alfonso, X. & Wiseman, I. (2007). *Local information systems: a review of their role, characteristics and benefits*. London: Department for Communities and Local Government.

Gunther, O. (1998). *Environmental Information Systems*. Berlin, Germany: Springer-Verlag.

Lincolnshire County Council. (2018). Lincolnshire Research Observatory. Retrieved from <http://shared.research-lincs.org.uk/Home.aspx>

NERC (2017). UK Soil Observatory. Retrieved from <http://www.ukso.org/home.html>

Nielsen, R.O., Chapman C.M., Louis W.R., Stovitz, S. D., Mansournia, M. A., Windt, J., ... & Finch, C. F. (2017). Seven sins when interpreting statistics in sports injury science. *Br J Sports Med*. DOI: 10.1136/bjsports-2017-098524

Scottish Government. (2017, November 03). The Scottish Index of Multiple Deprivation. Retrieved from <http://www.gov.scot/topics/statistics/SIMD>

Scottish Government. (2017, November 28). Access Our Data. Retrieved from <http://www.gov.scot/Topics/Statistics/About/DataAccess>

Scottish Public Health Observatory. (2014). The Scottish Public Health Observatory. Retrieved from <http://www.scotpho.org.uk/>

SEEMiS Group (u.d). *SEEMiS*. Retrieved from <https://www.seemis.gov.scot/site3/>

Skills Development Scotland (2017, August 29). The 2017 Annual Participation Measure for 16-19 year olds. The 3rd statistical publication reporting on learning, training and work activity of 16-19 year olds in Scotland. Retrieved from https://www.skillsdevelopmentscotland.co.uk/media/43580/2017_annual-participation-measure-report-29th-august-2017.pdf