

Scotland's Data Desert

The Case for a Scottish Statistics Agency

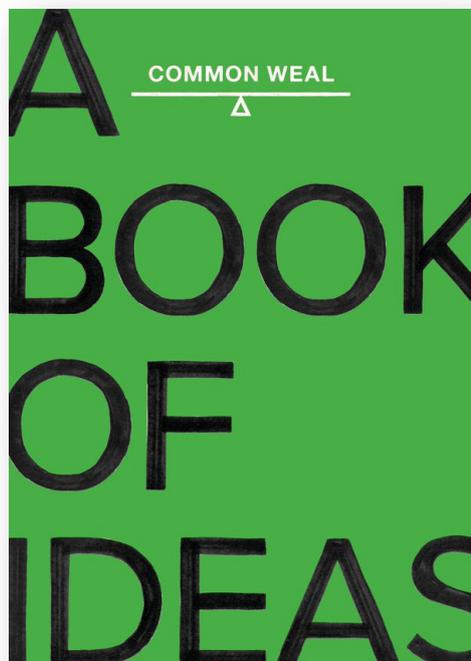
Craig Dalzell

February 2018

COMMON WEAL POLICY



COMMON WEAL is a think-and-do tank that advocates policies that put All of Us First. For more information on Common Weal Policy visit allofusfirst.org/policy or email ben@common.scot



Visit allofusfirst.org to get your copy of the Common Weal Book of Ideas 2016/21: 101 big ideas for the next term of the Scottish Parliament. Available to order or as an e-book.

COMMON WEAL
△

COMMON WEAL POLICY



Author

Dr. Craig Dalzell is the Head of Research for Common Weal. He holds an MSci and PhD in Laser Physics and Photonics and runs a politics and economics blog at thecommongreen.scot

Contents

Introduction

page 01

The Data Desert

page 01

Designing a Scottish Statistics Agency

page 05

Conclusion

page 10

References

page 11

Executive Summary

The ongoing discussions about Scotland's constitutional circumstances have highlighted significant shortcomings in the availability and quality of data regarding the measuring of many aspects of Scotland's economy, the demographics of its citizens and residents, data regarding its balance of trade with the rest of the UK and with the rest of the world. Whilst a case can be made that the running of the UK as a unitary state precludes the need for many such data, particularly that regarding Scotland's trade with the rest of the UK, the fact that Scotland now has substantial and growing devolved powers and the fact that such data is vital to discussion about the future of Scotland requires the creation of robust statistics. This paper shall outline several areas in which statistical data fails to meet required needs and shall make the case for the creation of a dedicated Scottish Statistics Agency tasked with finding solutions to this problem.

Key Points:

- Scotland's level of statistical data provision marks it as the best served nation within the UK.
- However, there remains substantial gaps in provision and overall provision falls short of what would be expected to serve an independent country.
- Regardless of Scotland's constitutional arrangements, better data provision is vital for the creation of better government policy and better management.
- Specific gaps have been identified in areas such as the regionalisation of data, the tax gap, GERS, Whole of (Scottish) Government Accounts, the National Register of Assets and Scottish import and export data.
- A Scottish Statistics Agency would be charged with identifying and filling gaps in statistical provision.
- The SSA may take the form of a single, centralised agency in which it is charged with collecting or commissioning the vast majority of statistical data.
- Alternatively it may be more of an overseer of a decentralised network of specialised data providers, a regulator to ensure the data meets a sufficiently stringent Code of Practice as well as maintaining a data portal to ensure access to the gathered data.
- Ideal budgeting is difficult to assess as they depend on the model employed. An SSA proportional in size to the UK's would have a budget of £15 million per year and would employ 265 people. An SSA the size of Sweden's would have a budget of £83.3 million and would employ 700 people.
- Where third party organisations are involved with gathering and manipulating data, a Code of Practice "kitemark" could be created which would indicate where such data meets the standards required for policy-making.
- Trust, transparency and ease of access to data are of paramount importance in this field and an ethos encouraging this should be built into the agency from the outset.

Introduction

The history of the gathering of economic data is deeply intertwined with the history of history itself. Some of the very oldest extant written documents produced by human civilisation – dating to the fourth millennium BCE – are records of stores of grain, animals and other logistics.¹

In the modern world, with interlinked economies spanning the globe and a level of complexity which is unprecedented, data gathering has become an ever more vital tool in the kit used by governments in their quest to create effective policy and to measure its impact. With the density of computational storage and speed of data processing being unsurpassed and still improving at a geometric rate it is becoming easier for governments to gather a wider scope of data at ever greater degrees of granularity. It has recently been remarked that data has now surpassed oil as the world's most valuable resource.²

Scotland already is both a consumer and producer of data for the purposes of policy-making and monitoring but it is often reliant on UK level statistics for this purpose. It has been described³ as a nation which is better served by data than the other constituent nations of the UK but nonetheless does not gather as much data as would be expected for an independent nation of its size and level of development.

Regardless of Scotland's constitutional arrangements, the issue of data is an important one as it should be universally recognised that better data can be used to produce better government policy and to better monitor policy as it is implemented.

It is time to discuss seriously the issue of data gathering and to look at the areas in which Scotland is still a relative data desert. The purpose of this paper is to open that discussion by identifying some of the key areas in which Scottish data gathering could be improved and to propose options for the size and structure of a Scottish Statistics Agency (SSA).

The Data Desert

As a region of the UK, Scotland is in many ways better served by data gathering and analysis than its counterparts. However, as Scotland takes greater control over its own affairs and as the constitutional debate continues to look towards a future in which Scotland takes full responsibility for its own affairs the question is raised of whether even this level of data provision is adequate for current or future needs – especially in a world where data becomes ever more vital in the development and support of policy. Well served though Scotland may be as a region, as a country it remains a relative “data desert” compared to other independent countries to which it could be compared. There are several areas of data which have been identified by this report and by others as omissions or areas containing significant limitations which could and should be addressed as a matter of importance. It should be noted

that this should not be considered a comprehensive list of known data problems in Scotland but is instead illustrative of the fact that issues can be identified across a broad scope of data gathering. These are also examples of specific limitations encountered by Common Weal in the course of policy research. It should also be stressed that these examples exist regardless of Scotland's constitutional position and can and should be addressed regardless of said constitutional circumstances. The fact that better data can be used to form better government policy should be a universally recognisable position.

Regionalisation of Data

Often statistics are gathered by means of representative samples. This means that a question or series of questions is asked for a small group within the country and their views are taken as a guide as to the views of the whole group. The precise techniques employed to perform such a sample have been well developed and studied over decades^{4,5}. In essence, so long as the sample is randomly selected and truly representative of the whole population (i.e. has the same proportion of people of various genders, ages or which ever other metric is important) then the expected deviation between the views of the sample and the views of the whole population should be governed only by the size of the sample (and not the size of the sample as a proportion of the whole). For a randomly sampled poll which has a sample size of 1,000 people this error should mean that 95% of surveys should give a result which lies within a margin of $\pm 3.1\%$ of the answer that one would get if the whole population was asked the same question. This should be true no matter how large that population is.

This assumption can break down when statistics are used to analyse specific sub-samples within them. A UK level survey of 1,000 people may be adequate for UK level decision-making but if that same sample is used to analyse the views of Scotland, policy-making becomes more difficult. In this case, one would expect only 83 Scottish data points within that 1,000 point sample (based on Scotland's population being around 8.3% of the UK's). Even a sample of 10,000 people across the UK (as is often used by organisations such as the ONS) would mean only 830 data samples in Scotland (and an expected error rate of at least 3.4%) which may be barely adequate for policy-level decisions. As sample size gets lower, the effect of any mismatch in proportionality within that sample becomes larger (one could imagine the difficulty of proportionately matching gender across all age and income bands) and the survey becomes more reliant on weighting and other mathematical modelling techniques to rebalance the survey. The practicalities of collecting survey data (for example, a face-to-face survey may incur significant travel costs when covering low-population density rural areas) can sometimes mean that samples themselves are not randomly selected. Techniques such as cluster sampling⁶ may result in UK-wide surveys which do not meaningfully disaggregate to a Scottish level despite seemingly large sample sizes.

Even if a UK level survey can be used for Scotland level policy, it is even harder to break down to local authority

level or even more locally than that. A hypothetical 10,000 person UK sample would imply an average of just 26 people selected from each of the 32 Scottish local authorities. If a Scottish local authority wished to study a UK poll of, for example, income distribution within its own boundaries then there would not be enough people to populate the desired income brackets in a meaningful way. There may even be issues with releasing the data to the local authority or non-government bodies at all as the low sample size could mean that individuals could be identified from the data thus data protection laws become important.

The trivial solution to this problem would be to commission more surveys but the immediate barrier to this is the cost both in monetary terms and in logistical terms. Expanding the hypothetical 10,000 survey to one for each of the four nations of the UK would at least quadruple costs whereas a bespoke survey for each of Scotland's local authorities would require 32 surveys. Even this costly solution merely passes the problem further down the scale of localisation as it may be important to interrogate such surveys at a council ward level or even within individual communities. Even if such widespread surveying becomes practical and affordable, any differences in methodology between each of the regions or changes in regional boundaries may lead to problems in cross-correlating, comparing or combining data between them.

An alternative solution, advocated by analytical studies such as the 2016 Bean Review⁷ is to employ the use of "administrative data" - that is, data held by the government for purposes other than statistics. For example, instead of commissioning a survey of income, it could become standard practice to access the income tax records of all workers directly to provide what is known as "near-census" resolution of income data.

Such data collection and access need not be limited to examples such as income. The Bean Review gives the example of VAT returns being matched with the business register to improve GDP estimates and other tax data used to improve collection of data from economic groups such as sole-traders which may not be captured by other statistical models.

In order to capture such administrative data in a statistically useful manner, it would require some level of co-ordination and forward thinking design – likely managed and regulated by the national statistics agency – so that each of the government departments involved could integrate their data with the needs of the statistics agency. Planned design formats would also ease the integration of new fields of data in the future as changes to technology, the economy and statistical modelling techniques may in the future allow the measurement of types of data currently deemed impractical to collect, not useful for current needs or not yet considered for collection at all.

The Limitations of GERS

Perhaps no feature of Scottish financial reports receives more attention than the annual Government Expenditure

and Revenue Scotland (GERS) report. This report lays out the annual revenue generated by the Scottish economy and the annual public expenditure attributed to Scotland – this lies distinct from the Scottish Government's annual budget which details devolved expenditure only and does not catalogue reserved expenditure either within or on behalf of Scotland and nor does it track revenue.

Whilst criticism of the use of these figures with regard to estimating the finances of an independent Scotland has been published elsewhere,⁸ the focus within this report will be to examine the structural weaknesses of this and related financial estimations with regard to the current status of a devolved Scotland within the United Kingdom along with the impacts of the currently evolving devolution framework.

The publication has been refined and revised multiple times^{9, 10, 11} since its inception in 1992. Given the continuing development of devolution in Scotland and the transfer of powers from UK Government to Scottish Government level it is likely to continue to be an active area of research and methodological development.

The evolution of devolution as well as distinctions in the methodology applied between the Scottish and UK governments has led to discrepancies between publications such as GERS compared to publications such as the UK Country and Region Analysis (CRA), which attempts to measure government expenditure in each of the nations and regions of the UK, and the UK Public Expenditure Statistical Analyses (PESA), which focuses on an overall UK view of public expenditure per government department. Additionally, a specific issue arises in the use of this category of data when referencing the potential finances of an independent Scotland. Both GERS and CRA apply a "who benefits" approach to expenditure such that expenditure which "benefits" the whole of the UK (such as military spending overseas) or which "benefits" Scotland solely but is the result of spending out with Scotland (such as reserved Civil Service work conducted from Whitehall) will be counted as Scottish expenditure, whereas tax revenue is largely counted on a "who pays" basis which attempts to count the revenue gathered from sources of tax located within Scotland. Analysts have noted^{12, 13} that this raises the possibility of tax revenue accruing due to government activity for - but out with - Scotland. Sources of this accrual may be both direct, for example, the income tax paid by a London-based Scotland Office civil servant, and indirect such as the rates and taxes paid by businesses frequented by said civil servant. It should be obvious, therefore, that if Scotland became an independent country and reserved government functions were moved from London to Scotland then additional income may accrue to Scotland's national accounts. Whilst this may be an important issue with regard to constitutional debate, it may not be so important for the purposes of financial analysis under the current constitutional framework.

Another related limitation on statistical publications such as GERS, CRA and PESA is revealed when one tries to cross-analyse the three publications. GERS and CRA both report expenditure in (or for) Scotland per public

spending category (for example, Scotland spent £927 million on “environment protection” in 2015-16)¹⁴ whilst PESA reports on spending by department on each of these public spending categories (HMRC spent £3 million on “environment protection” in 2015-16).¹⁵ However there currently exists no easily available or regularly published way to combine these datasets to find – as per this example – the amount spent by HMRC on environment protection in Scotland. Though some examples of this kind of missing data may appear trivial, they speak to a broader problem of the underlying data infrastructure not being able to adapt easily to cases where questions are asked which may not have been considered by the originators of the datasets.

Whole of Government Accounts

Whilst accounts such as GERS may allow a reasonable picture to be formed of the public finances of Scotland and accounts such as the Scottish Government’s annual budget allows a view of the annual direct spending from the Government, these two datasets are still limited and do not record the full overview of public sector revenue and expenditure in, for and by Scotland.

For this purpose, accounting in the form of whole of government accounts (WGA) is required. These reports – such as those published annually by the UK Government and with a UK outlook – record not just spending and revenue for each government department and publicly owned body but also fully list the assets and liabilities of these departments and bodies which includes accounts of debts owed to or owned by various departments or items, including records of pension liabilities and contingent liabilities.

An advantage of such an account is that if there is a transfer of liabilities from a publicly owned, yet operationally independent, body to the government’s own balance sheets (a hypothetical example could be a bank being nationalised then – some time later – its pension liabilities being taken on directly by the government) then if only the government budget is considered it may appear as though the government’s liabilities had increased at the point of the pension transfer whereas the WGA would reveal that it had merely transferred from one government body to another.

An inverted example may be in the possibility of the government moving liabilities “off-book” by setting up an arms length public body so that the government could borrow money for projects without inflating its own balance sheet or budget deficit.¹⁶ This was part of the rationale behind the Public Finance Initiative (PFI) scheme which saw the UK and Scottish governments take on comparatively more expensive finance deals than they were capable of obtaining through their normal borrowing powers with the rationale for doing so being at least partially because the political weight of taking on more visible debt was judged to be higher than the financial costs of the borrowing.¹⁷

It should be noted that Scottish financial publications like GERS have, in recent years, included some “off-book” financials like PFI as an aside and the Scottish Government

is, as of the time of writing, considering proposals to begin producing formal WGA accounts for Scotland. It would be critical that any eventual WGA produced should include the full range of government revenue, spending and liabilities and should record those liabilities at their fully discounted rate (i.e. including the final costs across the lifetime of the project rather than just the initial, upfront capital costs), as it has been reported¹⁸ that the UK WGA figures may currently be misstating such costs.

The National Register of Assets

The ongoing Scottish independence debate has focussed sharp attention on the issue of the division of debts and assets between Scotland and the remainder of the UK at the point of independence. The case has been made in a previous publication¹⁹ for a “zero-option” division of assets and debts consistent with the remainder of the UK adopting “continuing” or “successor” status to the UK and with Scotland adopting the position of being seen as a “new” state. Under this model, absent of agreement or negotiation to the contrary, public assets would be broadly declared as either fixed or moveable. Fixed assets include categories such as government buildings, military bases, roads and other structures, as well as mineral assets and land rights. These are almost always split geographically regardless of value with only a few exceptional circumstances to the contrary – often involving long term leases paid from the former state to the new one.²⁰ Moveable assets such as military equipment, publicly owned artworks and government vehicles, however, are often subject to sometimes intense debate over the precise division formula whether this is determined on the basis of need, population, territory, contribution or some combination of all factors.

With such weight being rightfully placed on this matter it seems surprising then that detailed assessments of the status and value of public assets either solely within Scotland or as part of a UK-wide assessment are not readily available.

The last UK wide National Asset Register was published in 2007²¹ and includes data recorded up until 2005. This study assigned to Scotland a total fixed asset value (excluding defence) of just under £23 billion – approximately £31 billion in 2017 prices²² – which is 6.8% of the total assigned to the UK. As stated above it can be expected that these fixed assets would simply be divided geographically. However, the National Asset Register does not contain reliable records of mobile assets and is therefore of limited utility for the purposes of forwarding the independence debate, calculating any expected share of UK resources or even for attempting to identify what essential assets an independent Scotland (or, for that matter, the remainder of the UK) may lack in the first instance should a simple geographic division be agreed.

For this primary reason it is an urgent recommendation of this paper that a Scottish National Asset Register be commissioned by the Scottish Government ahead of any future independence referendum so that negotiations over such assets can proceed as soon as possible after the

results are known. Further to this – likely once Scotland has become an independent country – it would be considered an important act of government transparency and accountability if regular audits could be published as it would, amongst other advantages, protect against the government of the day selling off such assets in a manner which could not easily be tracked.

Balance of Trade

Scotland's balance of trade, the difference in value between its imports and exports, is another statistic both extremely important to the economic management and policy making of the country as well as an issue often discussed with reference to the independence debate. In particular, Scotland's balance of trade is often mentioned in the context of its balance of trade with the remainder of the UK on the one hand and the balance of trade between Scotland and the rest of the world on the other.

There are two principle data sources from which Scotland's balance of trade is estimated. The first being the quarterly Index of Manufactured Exports (IME) produced by the Office of National Statistics and from which Scottish figures are extracted and extrapolated. This data is acknowledged by the Scottish Government as being less useful for estimating absolute value of Scottish exports and imports as it is based on modelled sub-sample data taken from the UK – its use within government is as a method of tracking the prevailing trends within the trade balance.

The second source is the annual Export Statistics Scotland (ESS) publication²³ which is based on the Global Connections Survey commissioned by the Scottish Government. This dataset is created not by direct measurements of goods at customs points but by surveying a representative sample of businesses based in Scotland which export goods outside of Scotland. However, the Survey lacks the statutory power to compel the return of surveys so the response rate can be particularly low, typically around 30%. This has implications for any study which attempts to break down figures to a sectoral or sub-national level as the sub-sample sizes can become too small for statistically significant analysis.

A 2007 analysis²⁴ produced by the then Scottish Executive outlined some of the limits of relying on such surveys as a source for trade data. Companies which trade from Scotland to the rest of the UK report that it can be difficult to identify when goods have been exported as goods are sometimes only tracked as being sold within the UK without reference to their location with respect to Scotland. Intra-UK trade may also be difficult to identify due to the complexities of supply chains within the UK whereby sales cannot be easily attributed to a single region – a particular concern of service sector firms or firms reliant on online sales and warehouse distribution. Finally, some sectors may simply be unable to track sales as an export or unaware that they may do so – a hotel room in Scotland which is let out to a non-Scottish resident constitutes a service export but this may not be regarded as such by the business replying to the survey.

One advantage of this method of data sampling, however, is that it eliminates discrepancies in export figures caused by export route. One example sometimes (wrongly) cited has been the “English ports” problem where it is claimed that Scottish goods which transited to a port in England before being exported could be counted as an English export and therefore omitted from Scottish figures. However, ESS stresses that because it asks exporters directly to state where they export to, only the start point and final destination are counted in the data without reference to the route taken by the goods.

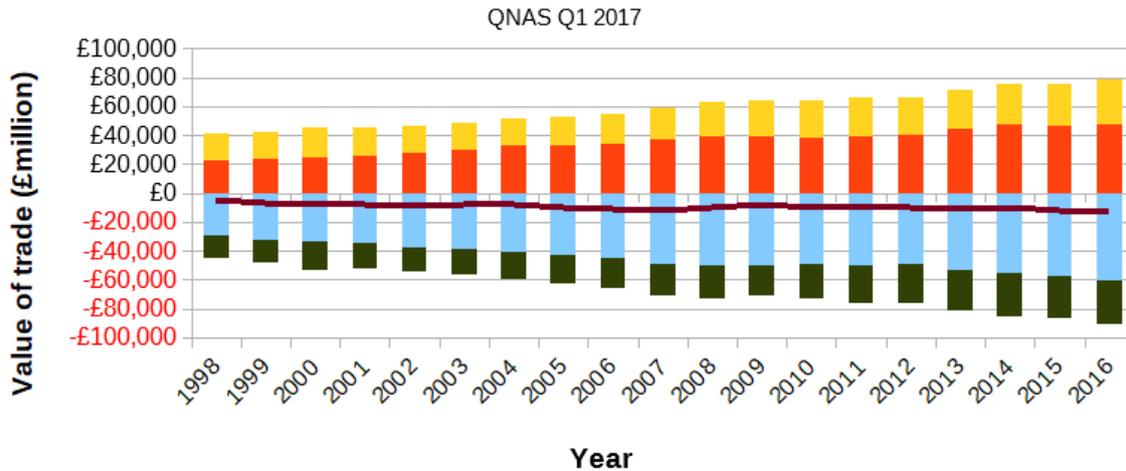
If the data on exports is problematic, the data on imports is practically non-existent. Data on imports into Scotland is not routinely or comprehensively collected nor is it currently possible to adequately track import flows in Scotland from the rest of the UK using UK level data. It is therefore not possible to directly measure Scotland's overall trade balance either with respect to the rest of the UK or with the world as a whole. Given the weight placed on balance of trade in many economic decisions and projections, this should be taken as a significant gap in Scotland's data collection.

At least part, and potentially a substantial portion, of the underlying cause of this lack of trade data appears to be due to the diminution of the role of the department of customs and excise in the UK. This department was merged with Inland Revenue in 2005 to form HMRC. Between the telescoping of capacity caused by this merger and the political decision to target immigration rather than trade, the ability for the UK to track exports and imports has been significantly compromised.²⁵ The restructuring scheduled for HMRC under the “Building Our Future” program is expected to result in the transition of all HMRC operations to 13 regional centres (2 of which will be in Scotland) and the closure of all other offices with the exception of five specialist centres. This is expected to lead to further degradation of the UK's ability to fulfil customs obligations at or near the point of entry or departure of goods.²⁶

Figures have been published recently by the ONS²⁷ which cast significant doubt on the validity of the data published on the balance of trade in services. The UK claimed to have a services trade surplus with the rest of the world of \$77 billion in 2014 but data held by each of the countries that the UK trades with (i.e. the sum of their respective recorded trade balances with the UK) indicated that the UK had an overall services trade deficit of \$39 billion in 2014. At the time of writing, a full analysis of the causes of these data asymmetries is ongoing and it is not yet possible to draw a conclusion on where the “true” value of the UK's services trade balance lies with respect to the two figures indicated in the report.

Despite these difficulties, the Scottish Government does attempt to produce some information on Scottish trade in publications such as the Quarterly National Accounts Scotland (QNAS). These publications draw largely on studies such as IME as well as sub-samples of UK data and apportionments and residuals drawn from UK GDP data to attempt to estimate trade flows between Scotland and the

Scotland Trade Balance



rest of the UK (RUK) and Scotland and the rest of the world (ROW) to attempt to estimate these trade flows.²⁸

Substantial caveats must be applied to figures such as this. The publication makes clear that the figures count economic activity in Scotland's offshore regions as RUK and that no estimates currently exist which allow disaggregation of this data. The publication also makes clear that the weakness of export data and the lack of direct import data mean that the figures should be treated with some caution and are liable to revision. Evidently a publication based on more direct measurements rather than statistical modelling of sub-samples will provide more robust data although it remains unclear how far movements towards this could proceed – in particular with regards to trade with RUK – within the confines of a unitary state.

Designing a Scottish Statistics Agency

Despite the limitations outlined in this report, Scotland is likely the most comprehensively served nation of the UK with regards to data – likely due to the decision by Scottish Government employees to seek out certain gaps in UK data provision or to amend UK figures in order to have them better reflect the needs or circumstances of Scotland. However, at least one analyst has described Scottish statisticians as “opportunistic, but relatively powerless, statistical scavengers”²⁹ which serves to highlight the limitations to the ambition of data provision in Scotland.

This said, Scotland significantly lags behind the data provision required for it to operate on par with several other independent nations of comparable size. Economies have grown ever more complex and interconnected in recent decades and the ability to gather and use data in decision making has grown alongside the computational power required to process it.

The case for the founding of a dedicated Scottish Statistics Agency is now well founded regardless of whether or not Scotland becomes an independent country. However, the vision for the structure of that agency is perhaps less well discussed. There is a debate to be had about whether the SSA adopts a centralised model whereby it is the primary producer of government statistics and data or whether a more decentralised approach is taken. Both of these models are certainly valid and countries find their own place on this spectrum. The UK as a whole tends somewhat more towards a centralised model compared to Scotland whereas countries like Ireland tend more towards a decentralised model. Scotland currently occupies a “hybrid” space where Scottish level statistics are produced via a loose network of decentralised bodies but where any of these bodies must derive data from a UK level source, they invariably look to the UK's centralised system for that data.

There is no absolute “correct” way to organise a statistics agency and all of these models have their advantages and disadvantages. It is, however, critical that the SSA is designed so that it is capable of producing statistics of a quality sufficient for government needs, as the government is very likely to be the agency's largest and most important consumer of data. If the SSA is incapable of addressing government requirements then it is likely that the government will turn towards outside sources such as academic researchers and industrial/commercial lobby groups for policy evidence. Without the SSA's ability to regulate such data, the potential arises for the quality of statistical provision to reduce.

A Centralised Model

In a centralised statistics model, the government will rely on one central statistics agency – perhaps paired with an independent regulatory agency – for the vast majority of its statistics and data requirements. Other bodies which produce statistics may exist – notable examples within the UK Government being the Office of Budget Responsibility and the Bank of England – however these bodies would tend to either be very strictly advisory or, especially in

the case of the central bank, tend to use their statistics internally rather than seek to influence others.

In the centralised model, there may also be third party bodies who gather and publish their own data and seek to lobby government. The key difference is that only the official statistics body can have their output designated as “official statistics” or, if their data is certified by the UKSA as meeting the Code of Practice³⁰, may qualify for a “kitemark” designating them as “National Statistics”. Through this approach, the public and government are offered assurance that the statistics produced meet the highest standards of rigour and may be confidently used for the purposes of policy-making.

The UK relies heavily on the Office of National Statistics (ONS) for a large part of the government’s statistics and this body is overseen and regulated by the UK Statistics Authority (UKSA). The UK does not operate an entirely centralised model, however, as it also maintains the Government Statistics Service (GSS) which also produces a substantial amount of data for government including work which qualifies for “national statistics” status.

The advantage of a centralised model is that it allows tighter control and focus for the agency and absolutely ensures compliance with the correct level of methodological rigour. The ability for data to be fit to a specific format makes cross-comparison across datasets easier – assuming the regulations are sufficient to demand that this be the case.

The primary disadvantage of the centralised model is that at some point a decision is made to gather or cease gathering a particular set of data and if only a small group of people are making those decisions then there is a risk that either data is not gathered and it is only retrospectively discovered to be useful – such as when the Scottish constitutional debate heightened the importance of comparative economic data between Scotland and the rest of the UK. Similarly, inertia within the department can mean that data continues to be gathered simply because it was gathered in previous years. One illustration of this can be found in Australia’s regular publication of data related to the Australian dollar trade weighted index. This index was previously used to define the peg between the Australian dollar and the currencies of Australia’s trading partners³¹ but since 1983 the dollar has floated freely. Whilst the data still serves or has taken on new purpose since then, its primary role has diminished.

If an independent Scotland was to adopt this model of data collection it would have to create a Statistics Agency which would have to have the capacity to take on the full scope of statistics already undertaken by the UK as well as any additional statistics required to fill gaps or expand scope as discussed earlier in this paper.

It would be within the powers of the devolved Scottish Government to create a new, centralised Scottish Statistics Agency and to draw in the same scope of data gathering as would be required in the case of an independent Scotland, but the cost of doing so may be higher as Scotland would,

presumably, still be required to pay a share towards the budget of the UK statistics bodies (For example, the UKSA’s budget for 2016 was £181 million³² of which a population share of 8.3% or £24 million could be attributed to Scotland). It would then have to independently fund the new SSA at an adequate level, potentially equalling or exceeding that spent by Scotland on UK level statistics. There would also be a risk of work being duplicated in the sense that the SSA could spend significant resources “re-running” studies already carried out from a UK level. This may be valuable in the respect that it would allow the SSA to study a larger sample size than would be covered by a UK sub-sample, but the overlap of coverage may be seen as inefficient. One solution would be to co-ordinate efforts with the UK to avoid such overlap, but this may also require co-ordination and harmonisation of methodologies to ensure that the datasets can be combined.

A further disadvantage to the centralised model is that it may not adequately take into account the data needs of individual local areas. The difficulties described earlier in this report to do with the localisation of data and the limits of sub-sampling apply just as much from the local authority level with respect to the Scotland level as they do from Scotland to the UK level. Local diversity is also a significant factor and a centralised system may find it difficult or deem it to be non-cost effective to run a study on a particular issue which may not apply outside of a specific area. Scottish local authorities already have the ability, at least where resource capacity exists, to collect local data in a way which best suits their own priorities or needs and it may be advantageous to encourage this rather than centralising data away from them.

A Decentralised Model

The alternative to the centralised model would be one in which a network of smaller, decentralised data gatherers with specialisations by data area and/or by geographic area is established. Scotland already does this to an extent, such that groups like those who produce Export Statistics Scotland work independently from groups such as those behind the Scottish Index of Multiple Deprivation.

In this model, there would still be a central Scottish Statistics Agency but its role would be somewhat more limited than in the centralised model. The SSA would gather and process data from the wider network and could make strategic decisions about what data would be needed in future and whether it should fill any gaps itself or should commission or encourage other groups to fill them.

It would also be the responsibility of the SSA to set regulations and methodological standards and to ensure that data can be cross-compared in a logical and user-friendly manner. To this end, the “kitemark” system could be expanded so that one mark would identify compliant data produced by approved government departments and another mark would identify data produced by third parties (such as think tanks or independent analysts) which also comply with the regulated code of practice.

This would open the field of data gathering and analysis up to a wider ecosystem of researchers, allow for better scrutiny and accountability of data and help encourage greater public trust in data disseminated by the government. As stated earlier, however, such outside provision must be coupled with stringent regulation so that only that data which meets the “kitemark” standards is used as evidence to support policy, otherwise there is the risk that the quality of data and statistics – and hence, the suitability of eventual policymaking – could be eroded.

The SSA would also be charged with aiding communication between groups. One particular criticism from within data professionals is that sometimes a department will gather a dataset and then not publish enough data to allow cross-comparison. This can lead either to lines of enquiry being cut as correlations between various factors cannot be drawn or it can lead to the duplication of work where vital data has to be gathered multiple times by disparate groups.

Similarly, the combination of datasets often has to be done independently by those who wish to make the comparisons. This runs a risk of the introduction of error or confusion which may compromise the validity of any research or conclusions which depend on the comparison. If the SSA acts as a repository for the data and a regulator of methodology then it may also be able to act as a processing body which could allow researchers to request datasets in a manner which means that they could automatically be combined and compared. One could consider systems such as those used by Eurostat, the OECD, the Gapminder project and others which allows users to set the parameters they wish (to give a hypothetical illustration consider requesting data on income distribution per local authority and a longitudinal time parameter), the SSA database would then retrieve the data automatically and present it graphically and as a raw datafile.

Designing such a “front-end” system must bear in mind several distinct audiences which may seek to use data from the SSA. Government departments, local authorities and the SSA itself may require full access to raw data including potentially sensitive or identifying data. Political parties, academics and think tanks may require enough detail to be able to recombine or analyse data in order to produce original research but it may not need access to individual data which could have implications for data protection. Finally, the lay public may wish to have access to displays of final conclusions and the methodologies underpinning them but may not wish to access much more than this.

An advantage of a more decentralised model lies in placing statisticians close to and responding to the needs of the government (which will likely be the statistics service’s major customer). This is particularly the case with bodies like the UK’s GSS in which statisticians are embedded directly within the relevant government departments. This close operation also, at least potentially, allows greater and easier access to administrative data as departmental bodies may be reluctant to share such data with outside sources.

Care should be taken in such a model however as individual departments are more vulnerable to political interference than a larger separate body (one could imagine a particularly controlling government minister ensuring that their department runs exactly according to their whims). This means that the issue of regulation and other oversight mechanisms must be particularly robust.

Decentralised models can also lend themselves to fragmentation and duplication of effort so require additional support from the overall governing body to ensure that this does not occur or that the effects of it are minimised.

Finally, there is the question of statistical resourcing, which could be at risk of adverse funding priorities which could limit the development of new statistics or even threaten to reduce provision. This risk could be mitigated by ensuring that funding for embedded statisticians is controlled by the overall statistics governing body rather than by individual departments.

Trust and usability

As important as it is to collect adequate amounts of data for policy development and monitoring, it is just as important that the public is able to trust that the data gathered is appropriate and is being used and presented in a manner which resists obfuscation or manipulation.

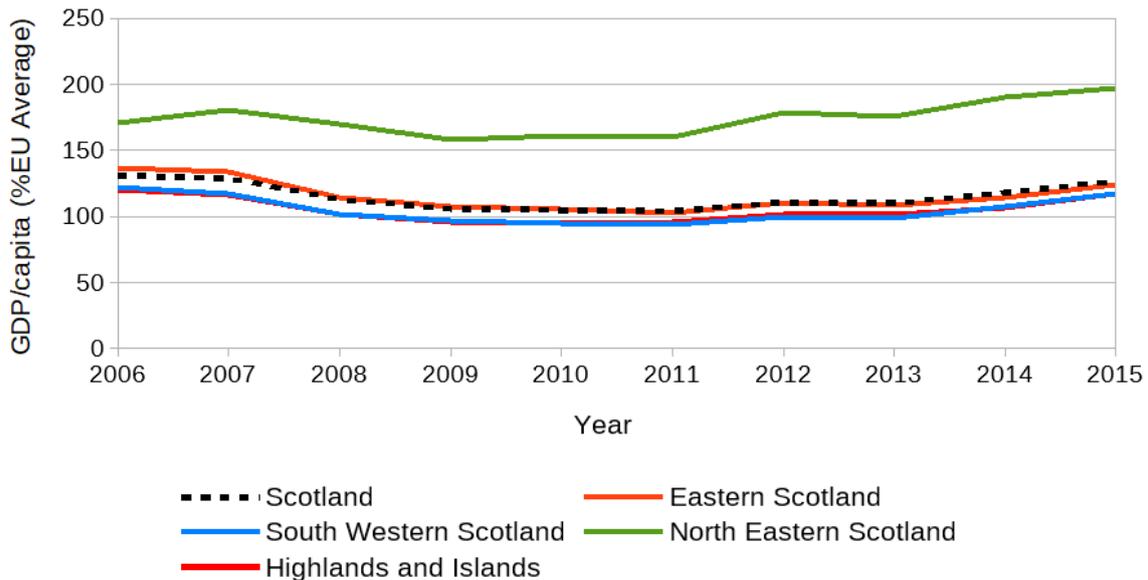
Trust and Transparency

The introduction of statistics to political debate has resulted in a general loss of trust among the public rather than the opposite. Complex figures can be thrown back-and-forth by both sides of a particular issue and outright misuse of data has not been uncommon. It is too easy to quote a particular number that supports a cause or agenda, whereas “fact-checking” that number can take orders of magnitude more effort. It should not be surprising therefore that scepticism of all data can become the default position.

Trust can be regained and maintained but to do so requires governments to be open about what data is being gathered, how it is being used and how it can be verified. The use of the data verification kitemarks mentioned above may play a role in this process.

Trust may also be maintained by building into policy and legislation resulting from or backed by statistical data a “who benefits” approach whereby a statement of the potential financial benefit of the policy is attached to the publication. Beneficiaries may be either direct in the case of payments to companies charged with delivering the policy or indirect from sectors which stand to gain from the decision. This would be particularly important in the case of a decentralised statistical system which employs non-governmental sources, as whilst it might be the case that a particular sector – for example, housing – has the expertise to deliver statistical information on the housing sector and are capable of doing it in a manner which meets the methodological standards set by the SSA, it would still remain the case that policy resulting in a boost to

Scotland's GDP per capita



the housing sector could directly benefit the body which produced that data, thus there is the potential for a conflict of interest.

Usability

The target consumers for data can have varied needs. Government bodies may need complete access to raw data – including potentially identifiable data points which would be covered by data protection laws if widely published. Academics, researchers and think-tanks may require sufficient level of detail of datasets to be able to re-combine and analyse them. Finally, members of the public may simply want to easily access the conclusions of studies and be able to clearly see how they were reached. To facilitate these needs, the Scottish Statistics Agency – whether it is based on a centralised or decentralised model – should build and maintain a single portal through which all kitemarked data can be accessed. A search and combination function based on setting search parameters can be designed which would allow users to request and download datasets as required. Eurostat, for example, allows users to request parameters such as “GDP/capita as a % of EU average” for each of the EU28’s NUTS-2 regions for a defined time period and allows the data to be downloaded in the user’s preferred spreadsheet format (See the graph below for an illustration of the data output of these parameters). The system could be designed such that users with varying access needs could be accommodated – for example, controlled levels of access to the site such that government agencies could request more detailed levels of data than members of the public.

Consistency and Methodology

As a substantial user of public data, Common Weal regularly comes across the issue of comprehensiveness and consistency. Without co-ordination from a central SSA, data can sometimes be gathered using different methodologies and cannot be easily combined. In Scotland this can particularly be the case with data gathering

initiated by local authorities, though data gathered by the Scottish Government can also differ in important ways from similar data gathered by the UK Government (for example, distinctions in tax revenue found in data gathered by GERS compared to HMRC³³). Data gathered by different methodologies can be difficult or impossible to cross-compare. Even when data has been gathered by similar methodologies, if it is stored in different formats (paper versus digital or even just differently formatted spreadsheets) it can be time consuming to re-format and the time and resources spent doing so often must be repeated by each individual or organisation doing the comparison, every time such data is accessed. This greatly reduces the efficiency of data analysis and greatly increases the chance of errors being introduced to the datasets.

A national SSA would have the ability not only to define a set of methodological standards but also a set of format and storage standards. It could even be possible to directly intervene and re-format data as required. This would be moderated by the abovementioned “kitemark” system as data which does not meet these standards would not be awarded the kitemark.

Transparency and Obstructionism

Almost everyone involved with public data will have had some experience of obstructionism in gaining access to that data. This can take the form of failure to publish, publication buried under non-intuitive websites (or published in hardcopy only with access only via appointment or similar controls), denial of freedom of information requests or use of “commercial confidentiality” or similar reasons to protect data from public scrutiny.

Commercial confidentiality during a tendering process is perfectly reasonable. However, once a contract has been issued there is little legitimate reason for people not to know how public money is being spent. If there are mild commercial disadvantages to a private company from clear public interest in knowing, the company must suffer the

disadvantage. Public contracts are lucrative and incredibly safe (this is a client guaranteed to pay its bills) and the price for this must be more transparency. The public is not wrong to suspect that the secrecy only benefits wealthy individuals.

Common Weal has noted the rise of obstructive practices especially in public procurement proceedings³⁴ and has recommend several proposals which would increase transparency and openness. These approaches should be built into the SSA from the outset.

Long Term Usability and Data Preservation

One major issue confronting users of data is the ability to use and preserve data in the long term. Some policies, such as pensions or surrounding healthcare, have consequences which will extend for decades beyond the initial announcement of the policy and it can be important to track outcomes for decades further. Older methods of data storage such as paper and velum are proven to be stable over centuries even in conditions of “benign neglect” (i.e. documents sitting in a box undisturbed) and examples of the oldest of all extant data storage media – the clay tablets mentioned in the opening section of this report – are still intact and readable many millennia after they were first recorded (although it should be noted that this was almost certainly not due to any deliberate intent on the part of the creators of said tablets). However, these storage methods are bulky and very difficult to search, extract data from or manipulate as can be attested to by many who have attempted to catalogue historical census, social security or local land registry data.

On the other end of the spectrum, modern electronic data storage can be quick and easy to search and manipulate (assuming the database is sufficiently well designed) but are subject to the whims of technological progress and the limits of material stability. A typical electronic storage device such as a hard drive or a CD-ROM may only be stable for a decade or so. It is also possible that technological change can render even once popular media completely obsolete. An infamous example of this was the BBC’s Domesday Project between 1984 and 1986 which carried out a wide ranging survey of census data, geographical and historical issues and other topics of interest within the United Kingdom. The data was stored on Laserdisc and designed to be viewed on the Acorn BBC Master Microcomputer, both of which became obsolete and fell out of production by the mid-1990s. It was only due to a substantial preservation program³⁵ that the data was saved and transferred into more modern formats in the early 2000s.

This is just one example of technological obsolescence affecting data preservation but it serves as an important warning. It may be that data deemed worthy of long term storage is preserved on a low volatility medium which can be readily copied onto more accessible but more volatile media as required (for example, data could be preserved on paper in a National Archive and digitised into a contemporary database at regular intervals or as electronic formats are updated). As data gathering becomes

even more sophisticated, one eye should be kept on the problem of long-term preservation. The economics of such storage (i.e. the costs of maintaining archives and updating electronic data stores) have been studied³⁶ but these issues are still open problems in the very long term.

Size and Budget

Given the highly variable nature of statistical provision between countries it is difficult to estimate the budget that a country the size of Scotland could be expected to spend on statistical provision. Countries employing decentralised models (or even hybrid models such as Scotland currently does) are difficult to audit as – ironically – total spending on statistics can be difficult to identify across all departments and not all departments can identify the total spend on statistics even within their own budgets.

Scotland does appear to manage a great deal of its most highly visible statistics work using very few staff. Just 0.8 Full Time Equivalent hours are used to collate the annual GERS report³⁷ and just 27 FTE jobs³⁸ cover Scottish economic statistics as a whole.

To give some illustration of an expected size of budget of a bespoke, centralised Scottish Statistics Agency, the example of the UK could be considered. The UK Statistics Agency has an annual budget of £181 million and employs around 3200 people³⁹. If an agency of a proportional size was created for Scotland it would imply an SSA with a budget of £15 million per year employing 265 people.

Other countries similar to Scotland also produce details of their own statistics agencies and spend significantly more than this projected “proportional share” budget.

Country	Annual Budget (£mn)	FTE Jobs Employed	Reference
UK (Proportional Share)	£15.0	265	[39]
Denmark	£45.5	530	[40]
New Zealand	£75.7	920	[41]
Sweden	£83.3	700	[42]

The comparatively modest costs of such an expansion of statistical provision in Scotland would be offset both by the economic benefit of providing these additional, highly skilled jobs and also by the economic benefits of the provision of the data allowing more efficient policy creation. As stated, a decentralised model may be harder to budget for and may draw data from statisticians outside of government, in which case the question of budget becomes rather convoluted so cost estimates involved would have to depend on the precise scope of the model adopted.

Conclusion

Whilst Scotland is reasonably well served by data when compared to other nations and regions of the United Kingdom, there is clearly scope for an order of magnitude of improvement. Gaps and weaknesses in the provision of data have now been identified independently by several sources and solutions have been offered.

An independent Scottish Statistics Agency with the statutory powers to compel data provision would reduce the impact of low sample return rates in surveys whilst a move towards greater use of direct administrative data would obviate the need for these surveys entirely – thus enabling a move towards “near-census” resolution of data provision across multiple target sectors. The Scottish Government already has most or all of the powers required to found such a Statistics Agency although arrangements may need to be agreed with regard to obtaining data from UK level government departments such as HMRC and the DWP.

A discussion should take place as to the level of centralisation or decentralisation involved in such an SSA as both models have their distinct advantages and disadvantages. The ability for non-governmental organisations to better provide credible and reliable data is a compelling argument towards a more decentralised model although the more focused strategic abilities of a centralised model may well be desirable.

If a decentralised model is tended towards then additional care must be taken with regards to the purpose and mandate of the governing body of the SSA. If statisticians are embedded within government departments then their independence must be protected from possible politicking within the department. If non-governmental sources are employed for data then standards absolutely must be ensured – perhaps via a “kitemark” system – so that the standard of provision is not degraded and the SSA must be able to ensure that the government is clear that only data which meets those standards is appropriate for use by the government.

Even in a centralised model, the SSA will likely be required to work with other statistics producing bodies especially whilst Scotland remains part of the UK and thus must co-operate and co-ordinate with the UK statistics bodies. This co-operation will, of course, run both ways and UK statistics bodies will likely have to be compelled to produce any data which impacts Scotland (whether it is Scotland specific or a sub-set of larger UK statistics) in such a way that it meets Scottish statistics standards and in a format which can be easily meshed with Scottish produced data.

In any case, a great deal of time should be employed in considering the nature of data accessibility as an open and transparent government is one which can be more easily trusted. A government which allows access to data is one which can be verified and one which can draw upon the expertise of specialist groups which may otherwise be locked out of decision-making.

Whilst this paper was motivated by the needs of an independent Scotland to build its own Scottish Statistics Agency, many of the challenges raised by the investigation can and should be addressed regardless of constitutional arrangements. Scotland requires data for government use under present circumstances and as devolution arrangement evolves the challenges inherent to data use evolve too. Indeed, there are challenges inherent to devolution – such as the calculation and analysis of the Scottish Government’s Fiscal Framework – which would not exist in an independent Scotland. For these reasons, the recommendation that Scotland assesses and reforms its statistical provision should be one which transcends the politics of the constitution, even if the final result will be shaped by it.

Compared to its potential, Scotland is a data desert in a world where data is rapidly becoming the most valuable resource available to governments. But with some investment, care and attention even deserts may bloom.

References

- ¹ J Friberg, "Numbers and Measures in the Earliest Written Records", *Scientific American*, Vol 250, No 2, February, (1984)
- ² The Economist, "The World's Most Valuable Resource is No Longer Oil, But Data", *The Economist*, May 6th, (2017)
- ³ J McLaren, "Economic Data Inquiry Response EDI-018", Scottish Government Economy Jobs and Fair Work Committee, September, (2017)
- ⁴ W G Cochran, "Sampling Techniques (3rd ed)", John Wiley & Sons, (1977)
- ⁵ J E Bartlett, J W Kotrlík and C C Higgins, "Organizational Research: Determining Appropriate Sample Size in Survey Research", *Information Technology, Learning and Performance Journal*, Vol 19, No 1, (2001)
- ⁶ W G Cochran, "Sampling Techniques (3rd ed)", John Wiley & Sons, pg 233, (1977)
- ⁷ C Bean, "Independent Review of UK Economic Statistics", UK Government Independent Report, March, (2016)
- ⁸ C Dalzell, "Beyond GERS", *Common Weal*, November, (2016)
- ⁹ J R Cuthbert and M Cuthbert, "A Critique of GERS: Government Expenditure and Revenue in Scotland", *Fraser of Allander Institute Quarterly Economic Commentary*, Vol 24, No 1, (1998)
- ¹⁰ J R Cuthbert and M Cuthbert, "GERS: Where Now?", *Scotland's Economic Future*, October, (2011)
- ¹¹ Scottish Government, "GERS Consultation 2016; Initial Response", July, (2016)
- ¹² C Dalzell, "Beyond GERS", *Common Weal*, November, (2016)
- ¹³ R Murphy, "Economic Data Inquiry Response EDI-003", Scottish Government Economy Jobs and Fair Work Committee, September, (2017)
- ¹⁴ Scottish Government, "Government Expenditure Revenue Scotland 2016-17", pg 32, August, (2017)
- ¹⁵ UK Government, "Public Expenditure Statistical Analysis 2016", pg 73, September, (2016)
- ¹⁶ G Donald, "Economy, Jobs and Fair Work Committee Official Report", Session 5, pg 32, September, (2017)
- ¹⁷ L Macfarlane, "Blueprint For A Scottish Investment Bank", *Common Weal and New Economics Foundation*, pg 10, October, (2016)
- ¹⁸ J R Cuthbert and M Cuthbert, "How to Have your Cake and Eat it: Government Accounting for PFI", *World Economics*, Vol 13, No 1, January (2012)
- ¹⁹ C Dalzell, "Claiming Scotland's Assets", September, *Common Weal*, (2016)
- ²⁰ J. Aitken, "Nazarbayev and the Making of Kazakhstan: From Communism to Capitalism", *Continuum Books*, p 166, (2009)
- ²¹ UK Government, "The National Asset Register", (2007)
- ²² Office of National Statistics, "UK Consumer Price Inflation", (2017)
- ²³ Scottish Government, "Global Connections Survey", (2014)
- ²⁴ C Munn, "Estimation of Scottish Exports to the Rest of the UK", *Scottish Executive*, July, (2007)
- ²⁵ B Austin and P Henderson, "An Independent Scotland's Customs and Borders", *Common Weal*, January, (2017)
- ²⁶ C Dalzell, B Wray et al, "Fighting For Tax Jobs, Fighting For Justice: A Workers' Alternative", *PCS Briefing Paper*, May, (2017)
- ²⁷ A Chesson, "Asymmetries in Trade Data - Diving Deeper into UK Bilateral Trade Data", *Office of National Statistics Release*, January, (2018)
- ²⁸ Scottish Government, "Quarterly National Accounts Scotland – Quarter 1 2017", *Economy and Labour Market Publication*, August, (2017)
- ²⁹ R Marsh, "Limitations and Strengths of Scottish Economic Statistics", *Scottish Government Economic Statistics Report*, March, (2017)
- ³⁰ UK Statistics Authority, "Code of Practice for Official Statistics", January, (2009)
- ³¹ C Dalzell, "Scottish Currency Options Post-Brexit", *Common Weal*, July (2016)
- ³² UK Statistical Authority, "Annual Report and Accounts 2015/16" pg 73, (2016)
- ³³ R Marsh and S Nicol, "Tax Revenue Estimates: A Comparison of GERS and HMRC", *Financial Scrutiny Unit Briefing*, October, (2013)
- ³⁴ M Cuthbert, "Public Procurement in Scotland", *Common Weal*, May, (2017)
- ³⁵ P Wheatley, "Digital Preservation and BBC Domesday", *Electronic Media Group*, June, (2004)
- ³⁶ D S H Rosenthal et al, "The Economics of Long-Term Digital Storage", *The Memory of the World in the Digital Age: Digitization and Preservation*, September, (2012)

³⁷ UKSA, “Statistics on Government Expenditure and Revenue Scotland”, Assessment Report 274, pg 5, February, (2014)

³⁸ Scottish Government, “Scottish Economic Statistics Plan 2017-18”, pg 21, August, (2017)

³⁹ UK Statistical Authority, “Annual Report and Accounts 2015/16” pg 61, (2016)

³⁹ UK Statistical Authority, “Annual Report and Accounts 2015/16” pg 61, (2016)

⁴⁰ S Emmel, “Introducing Statistics Denmark”, Briefing Note, (2014)

⁴¹ Statistics New Zealand, “Annual Report 2016”, pg 31, (2016)

⁴² Statistics Sweden, “Annual Report 2015”, pg 19, (2015)

www.allofusfirst.org

ben@common.scot

3rd Floor, 111 Union St,
Glasgow, G1 3TA

COMMON WEAL POLICY

