Return on Investment Factsheet

Effort: Hours

What is Return On Investment (ROI)?

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

ROI is a metric used to evaluate the merit of a single investment or to compare the relative merits of a number of different investments. It measures the amount of quantifiable benefit (return) relative to the investment's cost. To calculate ROI, the quantified benefit (return) is divided by the cost of the investment, and the result is expressed as a percentage or a ratio. A positive ROI means the benefits compare favourably to investment cost.

In business, the ROI metric is used to measure the rates of return and decide whether or not to undertake an investment. In government, ROI is increasingly used to compare and prioritize capital spending proposals within funding programmes. Within research, ROI metrics are often a feature of research infrastructure bids in the physical and life sciences but have been less often used for the humanities and social sciences.

ROI does not inherently account for the amount of time during which the investment is taking place. Hence one may also incorporate Net Present Value (NPV), a measure that accounts for differences in the value of money over time. For long-term investments, such as research data infrastructure where the benefits accumulate over several decades, the need for Net Present Value adjustment is high.

As a decision tool ROI is simple to understand. However you need to be aware of underlying variables and assumptions that affect the metric and how it was calculated. You can choose variables such as the length of the calculation time, or if overhead cost should be included, etc. To use ROI as an indicator to prioritize different investment projects is problematic unless the variables are defined and comparable.

The UK Economic and Social Data Service (ESDS) Impact Study and ROI

The ESDS impact study published in 2012 is currently the only example of a fully developed quantified economic impact study and ROI metrics for social science research infrastructure. It found that the quantifiable benefits and returns significantly exceeded the value of the funding invested in the ESDS. The study included two ways of expressing return on investment:

- There was a 5.4 to 1 benefit/cost ratio of net economic value to the service's operational costs;
- A counter-factual macro-economic approach based on returns to R&D at either 5% or 20% estimated the value of the additional re-use of the data hosted as £58 million to £233 million over 30 years (Net Present Value): suggesting a 2.5-fold to 10-fold return on investment.

Project funded by the EU Horizon 2020 Research and Innovation Programme under the agreement No. 674939

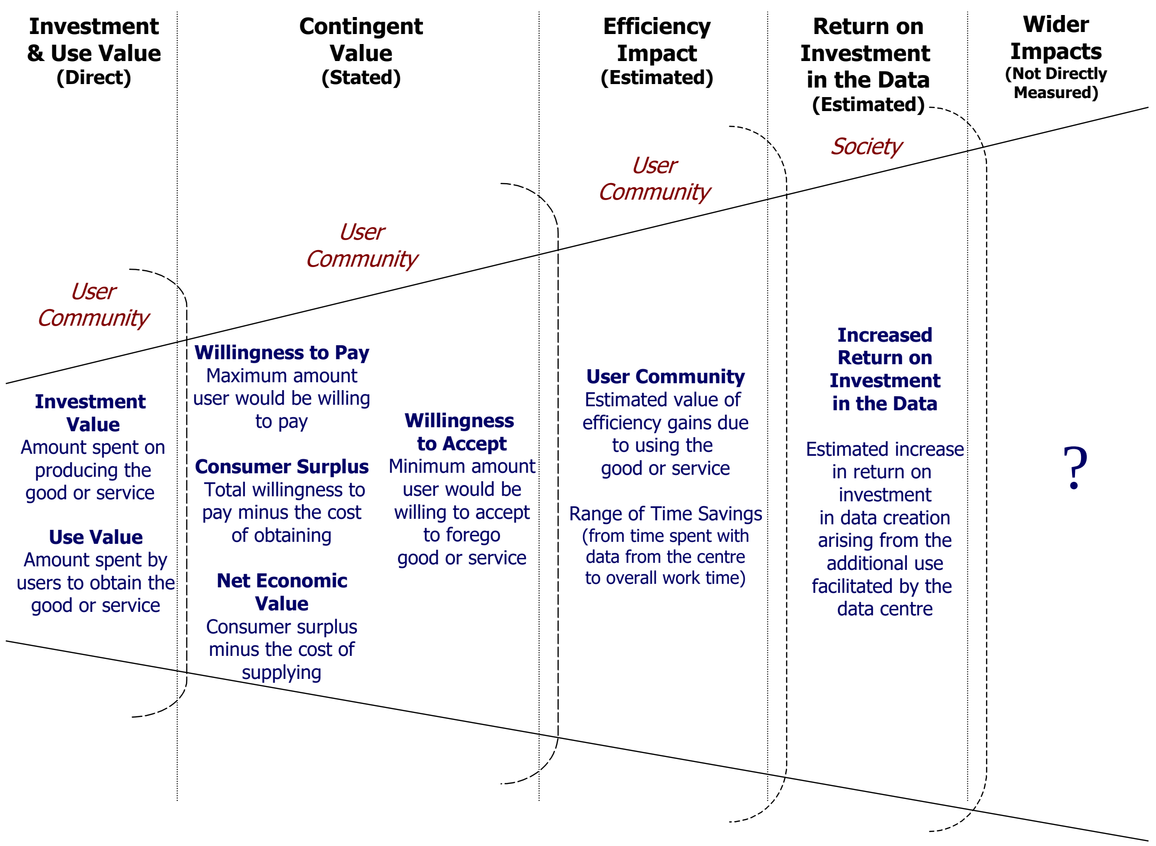
5.4 to 1 benefit/cost ratio

up to 10 to 1 returns on investment in data hosted arising from additional re-use

Illustration by Charles Beagrie Ltd ©2017. CC-BY licensed

Full details of the methodology used and underlying variables and assumptions used are available in the study report (Beagrie et al 2012). A short summary is provided below.

Methods used to Assess ESDS Impact and ROI



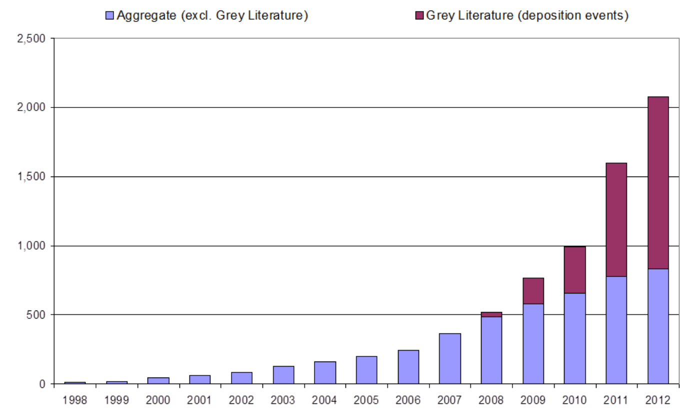
Methods for exploring the economic value and impacts of ESDS.
Beagrie and Houghton 2014 Figure 1. CC BY-NC Licensed

The economic methods used in the ESDS Impact study can be seen as estimating a range of values, moving from those focusing on minimum values, toward methods measuring some of the wider impacts. In parallel with this, a number of qualitative approaches were used including case studies, and the KRDS Benefits Framework as a method of presenting the broad spectrum of benefits analysed.

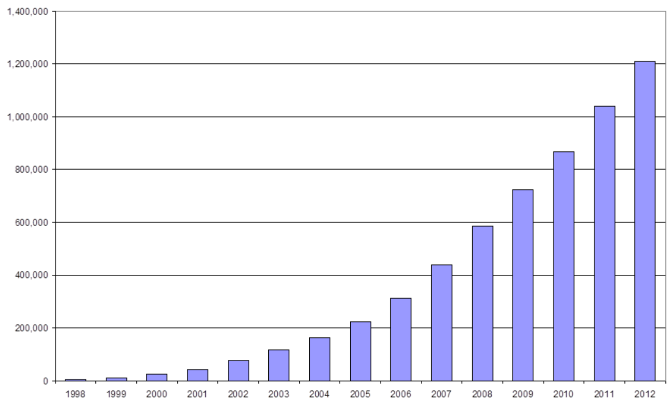
Maturity of services and collections and ROI

The ESDS was part of the UK Data Archive, a large mature social science archive. The approaches used and the findings on ROI in the ESDS impact study may be most relevant to similar archives. Less mature archives may need to consider how value and returns accumulate and can be measured over time.

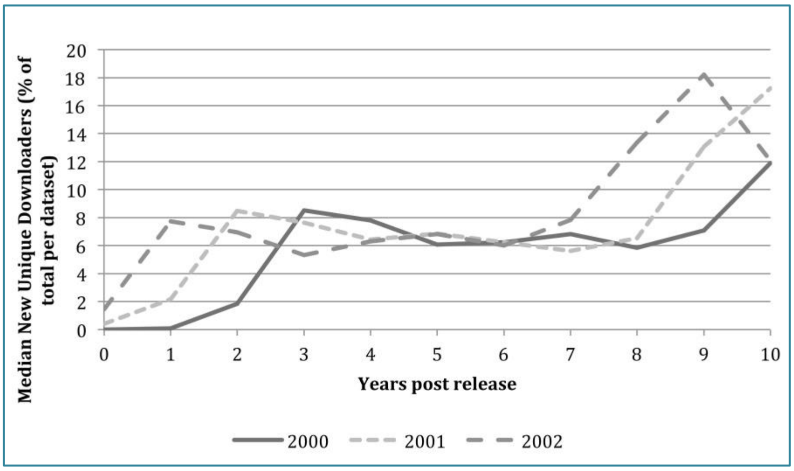
The ESDS impact study was the first of a series of impact studies on research data services in different disciplines (Beagrie and Houghton 2014). A unique feature of the Archaeology Data Service (ADS) Impact Study was the inclusion of an analysis of the evolving, cumulative value of the service. ADS provided historical data for their data deposits and uses and related operational costs back to 1997-8 (just over a year after its foundation in October 1996), which enabled the exploration of some indicators of cumulative value. That analysis indicated that the value of ADS data and services had increased as the collections had grown and the service had developed from its foundation (Beagrie and Houghton 2013). This is an important reminder that the maturity of collections and services at the date of evaluation may be an important factor to consider in future studies of less established data archives and ROI.



Archaeology Data Service collection deposits (deposition events), 1998-2012 (Beagrie and Houghton 2013, Figure 5.1)



Archaeology Data Service website accesses, 1998-2012 (Beagrie and Houghton 2013, Figure 5.2)



Relationship between time elapsed since ICPSR data for 2000- 2002 were released and the proportion of new unique downloaders gained per year (Fear 2013, Figure 5.5)

The impact of elapsed time on use (unique downloaders) of social science data sets was explored by Fear. Fear showed the relationship between the time elapsed since data sets were released by ICPSR and the proportion of new unique downloaders gained per year. The cohorts were time-shifted from one another but each annual cohort followed a distinct pattern: an increase in the number of unique downloaders gained per year to a rate that is maintained for about 5 years, then an additional increase. Overall, they showed a significant increase in use over a 10 year period from the point of release (Fear 2013).

Counter-factuals and ROI for new and emerging services

New and emerging services may face particular challenges with ROI because data collections are usually appreciating assets: returns can increase over time as collections reach a critical mass and user awareness of them grows. Similarly many investment costs are fixed costs and there is a minimum base for staff costs when starting up (see CESSDA SaW Costs Factsheet). Hence both initial returns and costs mean ROI for new and emerging services are likely to be negative for a number of years and then potentially have a strongly improving trend over time.

It may be helpful therefore to consider “the cost of inaction” and the counter-factual position if no archive exists. This can recognise that there are already hidden opportunity costs and depreciating negative ROI involved in doing nothing and provide a benchmark against which the value of new services can be assessed.

There are a small number of studies that have looked at quantifying some of these hidden and opportunity costs, particularly for data archived with individual researchers, in different disciplines and at different dates.

Vines et al requested the research data from a set of 516 articles containing morphological data from plants or animals that made use of a discriminant function analysis published between 1991 and 2011. They found that availability of the data was strongly affected by article age. For papers where the authors gave the status of their data, the odds of a data set being extant fell by 17% per year. In addition, the odds that they could find a working e-mail address for the first, last, or corresponding author fell by 7% per year. Their results reinforce the notion that, in the long term, research data cannot be reliably preserved and made accessible via individual researchers (Vines et al 2014).

Separately Vines et al examined how often requests led to them obtaining the data and examined the effectiveness of different approaches to data archiving. They contacted corresponding and senior authors up to 3 times over a 3-week period. Unlike the online data, which could generally be obtained within a few minutes, the requested datasets from individual researchers took a mean of 7.7 days to arrive. More than one e-mail had to be sent to the corresponding and/or senior author for 53% of papers, and the authors of 29% of the papers did not respond to any of their requests; no data were received 21 days after their initial request. They noted their average return of 59% in an average of 7.7 days is markedly better than has been reported in similar studies. They report requesting data directly from authors can also provide access to research data, but this approach is hampered by delays and the potential for disagreement between requester and the authors. Furthermore, the availability of datasets directly from authors will only decrease as time since publication increases. They concluded the results strongly emphasize the value of public databases for archiving scientific data and that money spent on data archiving in public databases is extremely cost effective (Vines et al 2013).

Pepe et al analyzed URL links for data embedded in papers published by the American Astronomical Society from 1997 until 2008 and their availability as of December 2011. Their analysis showed that the availability of linked data decayed with time. It also revealed that links to data hosted on astronomers' personal websites become unreachable much faster than links to datasets on curated institutional sites. In Figure 2 the average (arithmetic mean) rate of loss shown for links to data on personal websites is approximately 5.5% per year (Pepe et al 2014).

Wicherts et al attempted to obtain through e-mailed requests 249 data sets reported in 141 empirical articles in the last two 2004 issues of four major journals published by the American Psychological Association (APA). 6 months later, after writing more than 400 e-mails - and sending some corresponding authors detailed descriptions of their study aims, approvals of their ethical committee, signed assurances not to share data with others, and even their full resumes - they ended up with a meager 38 positive reactions and the actual data sets from 64 studies: 25.7% of the total number of 249 data sets (Wicherts et al 2006).

More recently, Krawczyk and Reuben report results of a field experiment in which two hundred e-mails were sent to authors of recent articles in economics that had promised to send the interested reader supplementary material, such as alternative econometric specifications, "upon request." Overall, 64% of the approached authors responded to their message, of which two thirds (44% of the entire sample) delivered the requested materials (Krawczyk and Reuben 2012).

Costs of Inaction: reported metrics for archiving via individual researchers			
Absolute loss	Rate of loss of research data sets	17% per annum	(Vines et al 2014)
Partial information loss	Rate of loss of working contact emails	7% per annum	(Vines et al 2014)
	Rate of loss for web-links to data on personal websites	c.5.5% per annum	(Pepe et al 2014)
Access	Data requests fulfilled	25.7% 44% 59%	(Wicherts et al 2006) (Krawczyk and Reuben 2012) (Vines et al 2013)
Delay	Elapsed time to fulfill data requests	Up to 6 months Within 1-3 weeks (mean 7.7 days)	(Wicherts et al 2006) (Vines et al 2013)

Although these reported metrics are from studies of different disciplines and study dates, they contrast sharply with the excellent preservation record, very high fulfilment rates, and rapid online access rates of public data archives in the social sciences. The public data archives also are appreciating as opposed to depreciating assets with improving rather than decreasing trends in value over time.

Using ROI in funding advocacy

One of the case studies in the toolkit looks at how the ESDS impact study published in 2012 (Beagrie et al 2012), has been used since in funding advocacy to Government by the service provider (the UK Data Archive) and its principal funder (the Economic and Social Research Council). It provides a valuable series of pointers to how ROI has been used and what has worked in funding advocacy in the UK, as well as potential transferable lessons for other social science archives and national differences to consider.

A lot of the focus when bidding to Government may be purely on numbers; easy ones to understand such as how many datasets, users, downloads, and so on. ROI is a simple metric that fits into these easily understandable and concise numbers. It can be particularly helpful to have when bidding in competition with other research infrastructures that already use ROI.

Data archives are appreciating rather than depreciating assets. Most of the economic impact is cumulative and it can grow in value over time, whereas most infrastructure (such as ships or buildings) has a declining value as it ages. Data becomes more valuable the longer you invest in its collection.

Therefore ROI is not an easy metric to calculate for data archives as economic returns are difficult to capture and can grow over time. However it is now been calculated for four mature data archives, from different disciplines, that have been in existence for at least 15 years (Beagrie and Houghton 2014, 2016).

These archives have also been of very different scale , complexity of services, and annual budgets (c. 700,000 euros – 56 million euros per annum). A significant investment of person-months and expertise is required to assess the ROI of a mature archive but the results may remain usable for up to 10 years and justify the initial investment.

If calculating a ROI for a specific data archive is not feasible, than citation of results from other studies such as ESDS can be helpful if they are relevant to your specific context.

For new and emerging archives, it may be more useful to focus on the "cost of inaction" and counter-factual metrics as ROI is likely to be low or negative in the early years.

Linked toolkit resources	Effort
Case study on the use of the ESDS economic impact study, http://dx.doi.org/10.18448/16.0005	hours
Costs Factsheet, http://dx.doi.org/10.18448/16.0003	hours
Benefits Factsheet, http://dx.doi.org/10.18448/16.0004	hours

Linked external resources	Effort
Beagrie, N., Houghton, J., Palaiologk, A., and Williams, P., 2012, *Economic Impact Evaluation of the Economic and Social Data Service*, http://www.esrc.ac.uk/files/research/evaluation-and-impact/economic-impact-evaluation-of-the-economic-and-social-data-service/	days

Other references

Beagrie, N. and Houghton, J, 2013, *The Value and Impact of the Archaeology Data Service: A study and methods for enhancing sustainability*, http://repository.jisc.ac.uk/5509/1/ADSReport_final.pdf

Beagrie, N. and Houghton, J., 2014, *The Value and Impact of Data Sharing and Curation: A synthesis of three recent studies of UK research data centres*, http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

Beagrie, N. and Houghton, J., 2016, *The Value and Impact of the European Bioinformatics Institute, Full Report*, http://www.beagrie.com/EBI-impact-report.pdf

Fear, K.M., 2013, *Measuring and Anticipating the Impact of Data Reuse*, University of Michigan PhD dissertation, https://deepblue.lib.umich.edu/handle/2027.42/102481

Pepe, A., Goodman, A., Muench, A., Crosas, M., Erdmann, C., 2014, How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers, *PLoS ONE* ,Volume 9 Issue 8: e104798, http://dx.doi.org/10.1371/journal.pone.0104798

Krawczyk, M., and Reuben,E., 2012, (Un) Available upon Request: Field Experiment on Researchers' Willingness to Share Supplementary Materials, *Accountability in Research: Policies and Quality Assurance*, Volume 19 Issue 3 pgs 175-186, http://www.tandfonline.com/doi/full/10.1080/08989621.2012.678688

Vines, T.H., Andrew R.L., Bock D.G., Franklin M.T., Gilbert K.J., Kane N.C., Moore J.-S.,. Moyers B.T, Renaut S., Rennison D.J., Veen T. and Yeaman S., 2013, Mandated data archiving greatly improves access to research data, *Journal of the Federation of American Societies for Experimental Biology*, Volume 27 pgs 1304–1308, http://www.fasebj.org/content/27/4/1304.full

Vines T.H. et al, 2014, The availability of research data declines rapidly with article age, *Current Biology*, Volume 24 Issue 1 pgs 94–97, http://dx.doi.org/10.1016/j.cub.2013.11.014

Wicherts, J.M., Borsboom, D., Kats, J., and Molenaar, D. ,2006, The poor availability of psychological research data for reanalysis, *American Psychologist*, Volume 61 Issue7 pgs726-8, http://dx.doi.org/10.1037/0003-066X.61.7.726