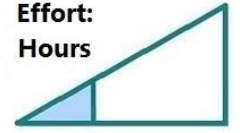


Case study on using benefit and cost tools Data Archive: Finnish Social Science Data Archive (FSD)

Effort:
Hours



Introduction



This case study examines how some existing benefit and cost tools could be used to determine the benefits of data archiving and the costs of this kind of research data infrastructure. Our focus is on the social science domain and on archiving survey and interview data. Preserving these data is particularly important because it is generally not possible, rarely economical, and sometimes not even ethical to replicate the data collection.

We are using the Finnish Social Science Data Archive (FSD) as our case study. FSD is a national resource centre that provides access to a wide range of digital research data for researchers, teachers and students. FSD is the Finnish Service Provider for CESSDA. Established in 1999, it has grown from 10 FTE to 24 FTE in 2016. From day one, FSD's key services have included data archiving, data dissemination and information services. FSD's data holdings contain 1300 studies and the Aila data download portal has 2800 registered users. FSD also provides support for research data management, participates in standards development and promotes open science. In 2014, FSD was awarded the Data Seal of Approval certification as one of the first CESSDA Service Providers. All in all, FSD can be characterised as a medium-sized data archive with a relatively high maturity level.

In this case study, we firstly take a look at the KRDS Benefits Analysis Toolkit. It is designed for use by a wide audience including data archives and repositories, and consists of two tools: the KRDS Benefits Framework and the Value-chain and Benefits Impact Tool. Secondly, we examine how to apply the ESDS economic impact study. Thirdly, we take a look at the CCEX Calculator. We aim to add insight to what is already known through previous research or from other components of the CESSDA Cost-Benefit Toolkit such as the Factsheets. This case study should therefore be read and used in conjunction with other components in the Toolkit.

This case study is likely to be of interest to all CESSDA Service Providers and other social science data archives, and their funders.

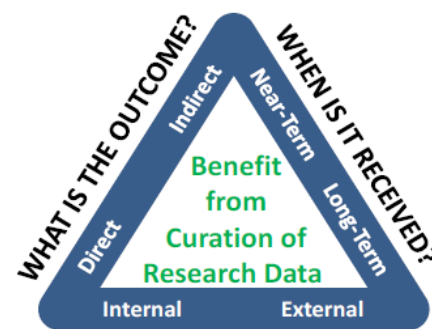
FSD © 2017. Image copyrights as attributed and reproduced with permissions. This work is licensed under a Creative Commons Attribution 4.0 International License (CC-BY). Requested attribution: Case study on using benefit and cost tools, FSD and CESSDA 2017, <https://dx.doi.org/10.5281/zenodo.3662456>



The KRDS Benefits Analysis Toolkit: Benefits Framework

The Benefits Framework is a tool for identifying, assessing, and communicating the benefits from investing resources in the curation or long-term preservation of research data. It organises benefits from the preservation of research data along three dimensions: what is the outcome, when is it achieved and who benefits from it (KRDS 2011).

The tool is an easy-to-use word template that provides a pick list of generic examples of benefits to help users to populate their own Benefits Framework. The *Benefits Summary for a Data Archive Worksheet* in the Toolkit and worked examples by FSD (see linked external tools below) and UKDS (Beagrie et al, 2012, p.43-44) provide helpful starting points for data archives.



WHO BENEFITS?
The Anatomy of a Benefit
(KRDS User Guide 2011 figure 10).
Illustration by Charles Beagrie Ltd ©2011. CC-BY
4.0 licensed

When filling in the template at the FSD, we found the following practices useful:

- Order benefits so that your most significant benefits in each section are at the top. They should be the first things a reader sees.
- In most cases, a benefit can be primarily associated with one category within a particular dimension. Choose the best fit instead of repeating a benefit in several sections.
- If necessary, extend the framework by subdividing the dimensions. For example, the time dimension could be divided into three: near-term, medium-term and long-term.
- Encourage everyone in your archive to contribute so that you can benefit from experts' diverse experience. Brainstorming techniques are helpful when gathering ideas.

Once you have compiled your Benefits Framework, don't let your work go to waste! Update and use it regularly. You may, for example, include it in your strategy planning process and use it for:

- Identifying strengths and weaknesses, and guiding which products and services should be included or excluded from the data archive's offerings
- Aiding and supporting discussion within data archives and with policy makers and funders
- Justifying data curation costs within funding applications
- Promoting data sharing

In this use case, we applied the Benefits Framework at the data archive level. However, it could also be used to organise and assess benefits associated with a particular activity or project within the data archive.

The KRDS Benefits Analysis Toolkit: Value-chain and Benefits Impact Tool

After the benefits have been identified and organised within the Benefits Framework, the next (but optional) stage is to identify potential measures or illustrations of the value and impact of those benefits. The Value-Chain and Benefits Impact Tool (VCBIAT) helps to identify quantitative metrics and qualitative indicators for the impact of benefits and supports a value-chain analysis.

The tool is a spreadsheet template that has been pre-populated with a selection of common benefits also used in the Benefits Framework Tool. You can choose to focus solely on identifying impacts from your benefits, in which case the tool should be relatively simple to use. You can also choose to add the research data lifecycle phases and activities in your analysis to see and evaluate a value-chain, in which case

familiarity with the KRDS Activity Model (KRDS 2010) is a benefit.

Whichever approach you choose, you will find that the VCBIAT is not as straightforward or easy to use as the Benefits Framework. It would be optimal to have a team of senior staff members to work with your VCBIAT. We would also recommend starting with a subset of benefits and/or lifecycle phases and activities; you can base your selection on your organisation’s most relevant needs and goals or even on your own expertise areas. Having said that, you should also keep in mind that benefits shouldn’t be evaluated in isolation as any action to realise a benefit may promote more than one generic benefit (Woollard 2011).

The table below shows a section of FSD’s VCBIAT focusing on the Data Sharing activity. As can be seen, the tool allows you to turn generic benefit statements into both quantifiable measures and qualitative interpretations of impacts.

KRDS Lifecycle Phase ⓘ	KRDS Activity ⓘ	Generic Benefit ⓘ	Your Expression of Benefit ⓘ	Action(s) to Realise Benefit ⓘ	KRDS Outcome Type ⓘ	Years to benefit ⓘ	Stakeholders who principally benefit ⓘ	Quantitative Impact(s) of benefit ⓘ	Qualitative Impact(s) of benefit ⓘ	Weighting of impact ⓘ	
Data Archiving and Sharing (Archive) ⓘ	Data Archiving and Sharing (Archive) ⓘ										
		Data Sharing (Access) ⓘ	Scholarly communication / access to data	Increased visibility / citations	Ensure easy and reliable access to metadata and data	Direct	Near	Data creators/owners; Data users	Number of citations to archived data increases; Number of data downloads increases	Data creators/owners gain scientific merit	3
			Re-purposing data for new audiences	Supporting multidisciplinary research	Ensure data is available for reuse (licenses in place)	Direct	Long	Data Users; Funders	Number of downloads from different organisations and disciplines increases	Breadth of data use increases; More research based on secondary analysis	3
			Re-purposing data for new audiences	Taking full advantage of existing data	Ensure data is available for reuse (licenses in place)	Indirect	Near	Data Users; Funders	Number of downloads from different groups (like students) increases	Breadth of data use increases; More research based on secondary analysis	3
		Long-term re-use of well curated data	No loss of future research opportunities	Ensure preservation processes are best practice	Indirect	Long	Everyone	Number of data downloads years after data collection	More data available for re-use; More opportunities for comparative research	4	

Value-Chain and Benefits Impact Tool (VCBIAT) analysis for FSD
FSD © 2017. CC-BY licensed

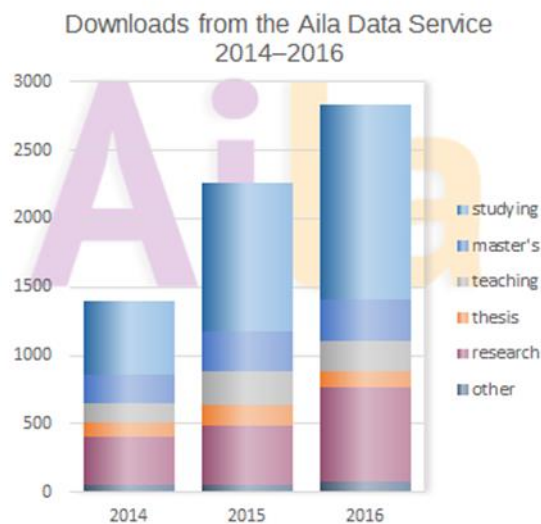
Just like the Benefits Framework, you can include your VCBIAT analysis in your strategy planning process. It can be used to assist evaluation, reporting, prioritising activities and maximising benefit and impact. It can also help understand the service impact your data archive has, explain the necessity of some the less visible activities and demonstrate value-for-money (Woollard 2011).

Economic impact evaluation: applying the ESDS economic impact study

The economic impact study of the Economic and Social Data Service (ESDS) quantifies the benefits and value of a social science research infrastructure using economic approaches. The key findings were that the quantifiable benefits significantly exceeded the value of the funding invested. There was a 5.4 to 1 benefit/cost ratio of Net Economic Value to the Service’s operational costs. A counter-factual approach estimated from 2.5 to 1 up to 10 to 1 return on investment in data and related infrastructure arising from additional use. (Beagrie et al 2012.) The study is discussed more fully elsewhere in the Toolkit: in the Return on Investment Factsheet, BenefitsFactsheet, and the UKDS Case Study on use of the ESDS economic impact study.

In 2010, FSD was awarded project funding to upgrade its services. One of the project goals was to improve

access to FSD’s data holdings by building an online data access portal. Since we wished to better understand the economic impact of our pre-ingest, ingest and access functions and to see the economic impact of the improved services, we decided to experiment with the method introduced in the ESDS study. We did not conduct a user impact survey or separate economic analysis, so we based many estimations and assumptions on the ESDS survey.



Downloads from the Aila Data Service 2014-2016

FSD©2016. CC-BY licensed

Two key parameters in the model are the number of accessed datasets and the number of registered users. At FSD, both numbers have grown significantly since we introduced our Aila data download portal in May 2014. As our fixed costs have remained the same, the model shows a substantial increase in FSD’s Net Economic Value, and also the estimated cost-benefit ratio has improved from negative in 2013 to 1.2 in 2016. We expect these figures to grow in the future as the number of our users and the number of downloaded studies are both growing steadily. However, FSD’s cost-benefit ratio is still far from ESDS’s 5.4, which suggests significant economies of scale and added user value from larger collections. It is also worth mentioning that FSD provides researchers extensive ingest services which increase our fixed (staff) costs; in the ESDS study, the researchers reported a mean of 185 hours preparation and deposit time,

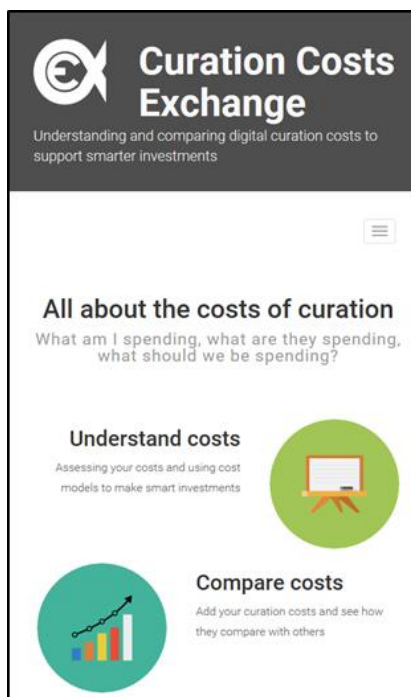
whereas we estimate that a researcher depositing data with FSD only uses 8 hours since FSD processes both data and metadata on behalf of the researcher.

In FSD’s case, the analysis was useful in showing that the investment in the data download portal has paid off and continues to provide economic value. We also gained better understanding of the costs and FSD’s economic value. Our analysis showed that fixed costs are significant for us, which aligns with the “rule of thumb” presented in the Costs Factsheet. However, there are limits to the extent to which the results of ESDS economic impact study can be utilised in a smaller-sized and/or less mature data archive.

Curation Costs Exchange (CCEX) Calculator

The Curation Costs Exchange (CCEX) is a tool that helps organisations assess the costs of curation practices through comparison and analysis. It was launched in 2014 by the Collaboration to Clarify the Costs of Curation (4C) Project. The project also produced an extensive evaluation of existing cost and benefit models, and analysed shortcomings in the capabilities of the models in relation to users’ needs for factual cost data and contextual information that defines these costs (Bøgvad Kejser et al 2014).

CCEX is an online tool. After registration, you can add information about your organisation and your cost data sets, and compare your costs with other organisations’ costs. The tool contains information about basic concepts used to assess curation costs and helps identify the cost drivers that may have an impact on investment decisions. In addition, it gives advice on planning for sustainability and recommends developing a business model using an approach like the Business Model Canvas (note the Archive Development Canvas tool in the CESSDA Cost-Benefit Advocacy Toolkit is based on the BMC).



The CCeX Website Home Page

The CCeX tool requires allocating costs to curation activities and attributing them to purchases (hardware, software, external services) and staff (producers, IT developers, operations, preservation specialists, manager).

At FSD, we used the CCeX tool to analyse the costs of pre-ingest and ingest, and data access. The tool itself was easy to use and if you have the required cost data at hand you can get results very quickly.

Since this kind of detailed information was not readily available from our systems, we ended up calculating crude estimations and making notes on how to record our cost data better in the future. The tool creates summary graphs of the costs and calculates relative costs in euros per gigabyte of data per year that can be compared with other organisations' costs. We did a global comparison and some peer comparisons, but unfortunately all other organisations were of "lower similarity" so the results were not that useful to us.

You are encouraged to share your cost data with others but it is not necessary. Due to the crudeness of our cost information, we chose not to publish FSD's cost data but instead used the tool for self-assessment only at this point. The CCeX tool can provide a good high level overview of your current cost structure. You can populate the tool with several years' cost data and analyse and predict changes in your costs structure. In addition, the possibility to compare your costs with similar organisations has a lot of potential. Knowledge acquired from using the CCeX can be used in strategic planning and decision making, and in developing your cost data records.

Key lessons for other data archives

Knowledge of the OAIS reference model is helpful when using any of the tools. The KRDS model and the CCeX tool both build on OAIS terminology.

The KRDS Benefits Framework is the optimal starting point. The effort needed is relatively small and there are good examples available. You can then proceed to any of the other tools presented here, depending on your needs and on your resources. The tools complement each other, each providing a slightly different view to costs and benefits.











Of these tools, most effort is needed for undertaking a study similar to the ESDS economic impact study. However, for each tool, the effort needed depends a lot on what kind of information you already have available.

All these tools can be used to help in strategy work, to justify a data archive's activities and costs, and to give visibility to data archives' "hidden" value-adding activities like competence sharing and standards development.

Peer comparisons help you gain insights from the experiences of others and recognise challenges. It would be useful if CESSDA service providers could agree on common practices in using cost tools, for example agree to define, qualify and structure cost data in a similar or comparable way.

Data archives are typically long-term investments and start-up costs as well as fixed costs can be

substantial. Size, scale of services, and maturity of the data archive are important factors to consider when using cost tools, interpreting results and comparing organisations.

Linked toolkit resources	Effort
<i>The Archive Development Canvas (Detailed Version)</i> , http://dx.doi.org/10.18448/16.0009	
<i>Costs Factsheet</i> , http://dx.doi.org/10.18448/16.0003	
<i>Return on Investment Factsheet</i> , http://dx.doi.org/10.18448/16.0002	
<i>Benefits Factsheet</i> , http://dx.doi.org/10.18448/16.0004	
<i>Benefits Summary for a Data Archive</i> , http://dx.doi.org/10.18448/16.0010	
<i>Case study on the use of the ESDS economic impact study</i> , http://dx.doi.org/10.18448/16.0005	
Linked external tools	Effort
Curation Costs Exchange (CCEX), http://www.curationexchange.org/	
KRDS Activity Model, http://www.beagrie.com/KRDS2_Activity_Model_detailed.doc	
KRDS Benefits Analysis Toolkit, http://beagrie.com/krds-i2s2.php	
Benefits Summary for the Finnish Social Science Data Archive, http://urn.fi/urn:nbn:fi:fsd:V-201703100001	

Other references
Beagrie, N., Houghton, J., Palaiologk, A., and Williams, P., 2012, <i>Economic Impact Evaluation of the Economic and Social Data Service</i> , http://www.esrc.ac.uk/files/research/evaluation-and-impact/economic-impact-evaluation-of-the-economic-and-social-data-service/
Bøgvad Kejser U., Hougard Edsen Johansen K., Thirifays A., Bo Nielsen A., Wang D., Strodl S., Miksa T., Davidson J., McCann P., Krupp J., and Tjalsma H., 2014, <i>D3.1 Evaluation of Cost Models and Needs & Gaps Analysis (Revision 1)</i> , http://4cproject.eu/component/docman/doc_download/55-d3-1-evaluation-of-cost-models-and-needs-gaps-analysis-revision-1
Consultative Committee for Space Data Systems, 2012, <i>Reference Model for an Open Archival Information System, CCSDS 650.0-M-2 (OAIS)</i> , http://public.ccsds.org/publications/archive/650x0m2.pdf
Woollard, Matthew, 2011, <i>KRDS Benefits Framework, Value-Chain and Benefit Analysis Tools: UK Data Archive case study</i> , http://www.beagrie.com/UKDA-KRDS_ImpactWorksheet.pdf