Guidelines for prospective data purchasers
## Contents

1. UK Data Service – Guidelines for prospective data purchasers 3
2. Scope 3
3. Background 3
4. Assessment questions 3
   4.1. Initial questions 3
   4.2. Secondary Questions 4
      4.2.1. Value for Money 4
      4.2.2. Data producer 4
      4.2.3. Payments 5
      4.2.4. Data quality 5
      4.2.5. Data format 5
      4.2.6. Updates 5
      4.2.7. Rights issues 6
      4.2.8. Support 6
      4.2.9. Availability 6
      4.2.10. Access control 6
      4.2.11. Back up and security 7
      4.2.12. Recurrent costs 7
5. Definitions 7

References 8
1. UK Data Service – Guidelines for prospective data purchasers

Public version
28 January 2015
Version: 02.00

2. Scope

These Guidelines set out a generic set of guiding principles and questions to ask about planned data purchase for research, typically by Research Centres. They should be read in conjunction with the UK Data Service’s Collections Development Policy (CDP) and Collections Development Selection and Appraisal Criteria which provide an overview of the selection and appraisal criteria applied to the UK Data Service data holdings. The successful implementation of the CDP ensures that only the most relevant and highest impact data collections are selected for ingest and curation; it also ensures that limited resources are not expended on ingesting potentially low-use data collections. Furthermore, the ESRC provides a detailed framework for research ethics to which it expects all research it support to subscribe.

Data considered for purchase require additional assessment questions to be addressed and these are not explicitly covered in the referenced CDP or appraisal documents. Neither is specific advice on the purchase of data covered, so privacy concerns can be more complex for some data sources, where explicit consent for use may not have been provided. The Collections Development Selection and Appraisal Criteria contain an Appraisal Grid used by the UK Data Service, which can be adopted and adapted by Research Centres to suit their specific needs.

The intended outcomes from this advice are for: the data purchaser to spend as little time as necessary on contractual and legal issues; the primary funder, in this case, the Economic and Social Research Council (ESRC), to ensure a relevant level of accountability; and compliance with the ESRC’s Research Data Policy. Where possible, compliance includes curation and management of data with the UK Data Service for the longer-term to maximise value for money in any purchase.

Some definitions are provided at the end of the document.

3. Background

The guidelines were initially developed to support the ESRC’s Big Data Network Phase 2 Business and Local Government Data Research Centres. However, they should be seen as a more generic set of guiding principles that set out the questions to ask about planned data purchase for research, including new and novel forms of data.

All of the guidance on the UK Data Service website relating to Data Management Planning remains relevant. This document should only be taken to cover those things over and above the standard Data Management Planning advice and guidance.

4. Assessment questions

4.1. Initial questions

Are these data acquired for a specific time limited research project?

If so, the justification for value may need to be more strongly stated.
Are these data being acquired for wider use?

Note: At present we foresee three scenarios. 1) Made available by Research Centres acquiring data to a wider user group for the length of an award; 2) Made available by the UK Data Service before the end of the award; or 3) Made available by the UK Data Service after the end of the award.

If the answer to the second question is yes, a more stringent process surrounding acquisition needs to be upheld.

Note: the UK Data Service is not the arbiter of the ‘relevance’ of the data to the original research programme proposed by Research Centres. If data purchase was proposed in the original research proposal that should be considered sufficient justification.

Is the supplier the sole supplier of these data?

Note: Some commercial data are re-sold by multiple organisations. ESRC, Research Organisation or local procurement rules may need to be followed. In the event of multiple data providers there might be procurement issues, depending on the cost. Even with a single supplier, a single tender waiver may be required. It will be up to the Principal Investigator to get advice from his/her procurement office.

Are there any indications of current or potential user demand?

Is there known user demand for the data beyond the Research Centre?

If there isn’t, is there a clear case for making data available beyond the lifetime of the Research Centre?

Does the data fill a significant gap in the existing data provision?

If it does, then there may be a clear case for making the data available beyond the lifetime of the Research Centre; although this question is more likely to be of value in other contexts.

4.2. Secondary Questions

4.2.1. Value for Money

Are there any measures which could indicate value for money?

Note: as a rule of thumb, for data being supplied for a time-limited research project, there is a case of paying the provider for the marginal cost of data supply, as they will also be benefiting from the impact of academic use. For data being supplied for wider use, commercial data providers should be compensated for loss of sales, i.e. they shouldn’t be made worse off by agreeing to a central purchase than compared to what they would normally expect to receive through academic data purchases. However, the additional exposure by academic users may also be ‘costed’ in.

Is the product of general relevance to the research goals of the project or of relevance to the ESRC’s priorities?

4.2.2. Data producer

Is there any reason to assume that the vendor/data supplier is not reliable?

Note: by reliable, we mean to make the prospective purchaser think about the likelihood of the supplier meeting any requirements which may be necessary within any contract, including being in existence before the termination of any contract. Note: the UK Data Service prefers contracts to be between entities which have a relatively low risk of disappearance, since curation activities depend on the rights owners of any data permitting the distributor/curator to copy and distribute.
the data lawfully. In the event of a company going bankrupt this right may disappear and would need to be dealt with contractually.

*Are there any conflicts of interest between the supplier and the purchaser?*

Standard ESRC guidelines on such conflicts of interest should be adhered to.

*Does the data producer require reports on usage?*

This should be established because if data are to be made openly available, for example without registration, then detailed usage (who and why) will be difficult to report.

### 4.2.3. Payments

*Is the payment for data “one off” for the duration of the project/perpetual? Are annual subscriptions required?*

Note: Terms of payment should be fixed at negotiation stage.

### 4.2.4. Data quality

*Has the quality of the data been assessed? Is there any reason to suspect that it is not “research quality”?*

*Is there adequate metadata and documentation for the data to be used by the Research Centre or other users?*

Note: As noted above the relevance of these questions may depend on the type of data. The primary aim is to ensure, in general, that the data has been assessed in such a way as to ensure that it is fit for the purposes for which it is intended to be used. The aim of this is to protect the investment of the ESRC; it is not supposed to stifle innovation. In some cases, these questions will have to be considered very carefully, in others, where exploration of the potential research value of the data are part of the research itself, it would not be necessary.

Note: depending on the circumstances, the UK Data Service may have to ingest these data into their collections. If adequate metadata and/or documentation do not exist, it may be possible that the UK Data Service is unable to ingest these data.

### 4.2.5. Data format

*What formats are the data to be supplied in?*

Principal Investigators should check which particular standards, if any, are supported e.g. DDI or SDMX.

*Will work be required to convert them to more usable formats?*

*Will work be required to convert them to appropriate formats for curation?*

*If there is work to be completed in the ingest process, are there the required funds to do so?*

Note: these issues may affect any decisions to be made on the value of the purchase. If the Research Centre is responsible for making data available, will the purchased data be delivered in formats which ensure that the data are usable by others without additional work? The same principle applies to data which may be supplied to the UK Data Service in a number of years.

### 4.2.6. Updates

*Is this a one-time purchase or will updates be required?*
Do the data represent a single momentary extract from a continuously updated database or one edition of a periodically updated dataset?

For how long are updates part of the original purchase price?

Is there a data supplier requirement to maintain access to ‘previous versions’?

Note: the level of relevance of these questions will depend on the types of data. If updates/new editions are made available, there should be clarity on the methods of version control of data files. Version control includes making sure that if the information in one file is altered, the related information in other files is also adapted, as well as keeping track of versions of data files and their locations. In general these principles allow the distributor to apply persistent identifiers to the data for research integrity and impact assessment.

4.2.7. Rights issues

Is there any third-party content in the data?

If so, the data provider will have to obtain clearance to include that content in any data made available to users.

Is this a purchase (indefinite use) or a fixed term licence to use data?

Note: If a purchase is for indefinite use, would it make sense for the ESRC or the University of Essex (as the legal entity behind the UK Data Service) to be the initial licensee?

Note: For data which is for indefinite use, there are many possibilities. The main concern from the point of view of the UK Data Service, is that transferring licences is often more complex than negotiating one in the first place. Thus, it will make sense to ensure that any licence agreement entered into with a data supplier takes into account the long-term needs.

Does the licence cover all the updates to the data?

Will the licence with the data owner explicitly provide for any relevant rights in any data derived from the original data?

4.2.8. Support

Where access is being provided, is there adequate support for the data product?

4.2.9. Availability

Are the data already available to the Higher Education community, but in a format which is difficult to use?

Note: this is unlikely to be relevant to most Research Centres, but if it is appropriate, it will help make a case for financial support.

4.2.10. Access control

What are the restrictions on the use of the data?

Are these restrictions appropriate to the sensitivity of the data?

If data are sensitive, who is responsible for output checking and disclosure control?

Are there explicit rules for output checking?

Are the licensees aware of the legal consequences of disclosure?

How will data be accessible?

Under what type of licence will the data be accessible?
Will the original data supplier need to be involved in the ongoing approvals of access or publish research results from the data?

Is there a requirement for any usage/performance statistics?

Note: these questions are designed to determine whether the types of access mechanism are appropriate for the data to maximise their use. They are also there to ensure that thorough consideration has been given to the legal and contractual responsibilities which are likely to affect the data purchasers, data users as well as the UK Data Service in time.

Are there any ramifications of the data being made open by the data supplier at some point in the future?

Note: the UK Data Service’s experience with international data providers (mostly providing aggregate data) is that intergovernmental organisations (IGO) may, during the lifetime of a contract, make their data openly available. In these cases partial refunds of the licensing cost are built into the contract.

4.2.11. Back up and security

Will the data and any metadata be backed up? What requirements are in use for information security?

4.2.12. Recurrent costs

Are there any likely recurrent costs attached to the curation of these data?

5. Definitions

Acquisition: The process related to acquiring materials for inclusion in the collection.

Archiving: A curation activity which ensures that data is properly selected, stored, can be accessed and that its logical and physical integrity is maintained over time, including security and authenticity.6

Curation: The activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose. Higher levels of curation will also involve maintaining links with annotation and with other published materials.7

Data Owner: The rights holder in any data collection. Not to be confused with the Data Producer.

Data Producer: An individual involved in the creation of a data collection, who may also be the Data Owner.

Data Supplier: An organisation which sells data. The supplier may also be the owner and/or the producer, but not necessarily so.

Data User: In this document, this term refers to those people accessing data under a purchase agreement. Likely to include: a) staff employed directly in relevant research activities; b) staff within Higher Education beyond the individual Research Centres; c) researchers in a commercial organisation; d) the public.

Depositor: A depositor is an individual who is named on a Licence Agreement as having sufficient responsibility to grant particular rights to the UK Data Archive on behalf of a data collection. The depositor may be the instigator, creator or the copyright owner of a data collection, but does not have to be. In OAIS terminology, the term producer is used in a similar sense.

Ingest: The processes related to receiving information from an external source and preparing it for archival storage and access.
Guidelines for prospective data purchasers

License Agreement: Used in this document as a generic term covering a range of different licences between an institution and the Data Owner.

Output checking: Vetting of research outputs by experts for disclosure risk.

Preservation: An activity within archiving in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology. Note that preservation differs from archiving in as much as preservation does not include maintaining the logical and physical integrity of a data object.⁸

Selection: The process of identifying materials to be made accessible and in many cases to be preserved for long-term use.

References

[URLs checked 25 January 2015]


⁴ ESRC Big Data Network Phase 2. http://www.esrc.ac.uk/research/major-investments/Big-Data/BDN-Phase2.aspx

⁵ UK Data Service Prepare and Manage Data. http://ukdataservice.ac.uk/manage-data.aspx

