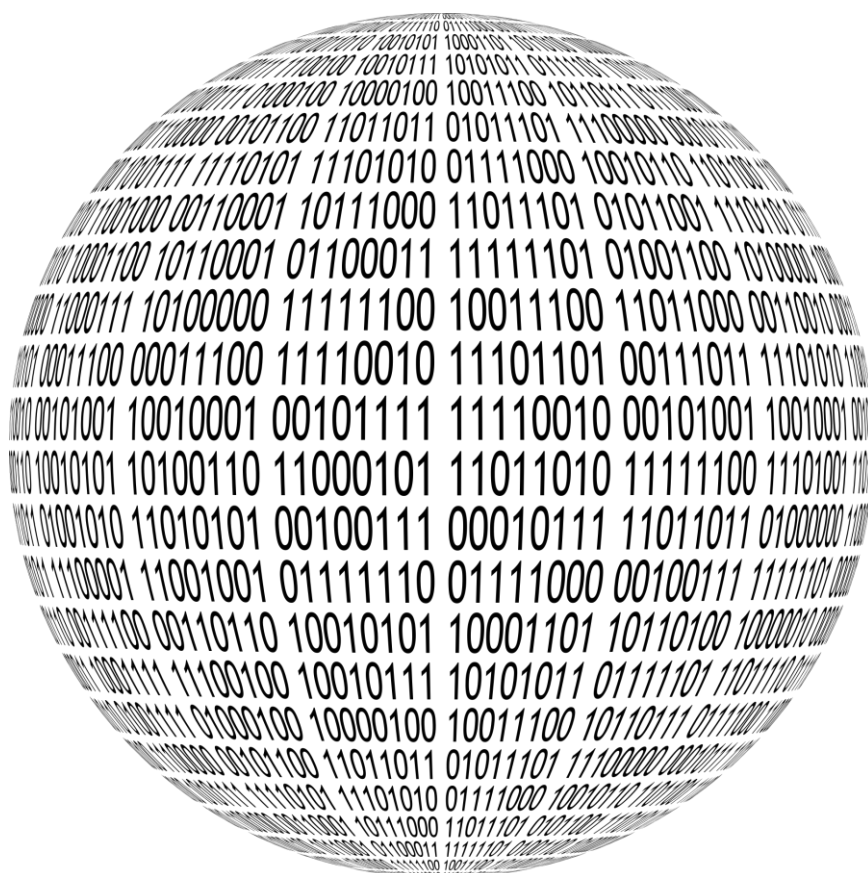


Data publishing case study: from collection to publication

SPHERE

Version 1.5 March 2023



University of Bristol

Research Data Service

INTRODUCTION

This document forms part of a series of data publishing case studies, each following a different project or research group in the Faculty of Engineering from initial data collection through to publication of findings. These case studies aim to showcase the ways that data sharing has been achieved across different projects with a variety of data formats and access restriction requirements.

Why publish data?

In this example, the authors are reporting work funded by EPSRC and are therefore expected to share their data as far as possible within ethical and/or commercial restraints. In addition, sharing data can increase the impact of a project, increase opportunities for collaboration, and thus maximise the benefit of public funding.

EPSRC guidance states that researchers should ensure that data underpinning published findings is available for scrutiny by others¹; however, what constitutes ‘supporting data’ is not defined. The following examples illustrate how researchers in the [SPHERE](#) project have interpreted this guidance.

EXAMPLE

Whitehouse S, Yordanova K, Paiement A, Mirmehdi M. “Recognition of unscripted kitchen activities and eating behaviour for health monitoring”. 2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2016). [doi: 10.1049/ic.2016.0050](#)

Dataset DOI: [10.5523/bris.raqa2qzai45z15b4n0za94toi](#)

Data collection

In this study, participants were asked to prepare a meal and/or drink in the SPHERE kitchen, which contains various environmental sensors and a camera recording RGB-D (depth) image data. In addition, participants wore head-mounted cameras. Working files were stored in the SPHERE-IRC project space in the Research Data Storage Facility (RDSF).

An ontology was developed to classify the various actions performed by the study participants, and this ontology was used to annotate the head camera data and to develop a model for kitchen activities. Finally, the model was tested against simulated data.

Selection of data for publication

Some data were unusable due to errors with camera positioning; the authors have excluded these data from the final published dataset. The phase of the study using environmental sensor data and bounding box data (derived from RGB-D images) is not yet complete, and has not been described in the current

¹ Clarifications of EPSRC expectations on research data management (UKRI) <https://www.ukri.org/about-us/epsrc/our-policies-and-standards/policy-framework-on-research-data/expectations/#contents-list>

publication, so the authors have not included these data in the published dataset. They have chosen instead to publish all usable data discussed in the paper, including annotated head camera footage, ontologies, and simulated data.

The study participants consented to share their data for research purposes only, so the dataset requires limitations on access. The data.bris repository provides a number of different access levels; in this situation the researchers have selected ‘restricted access’ as the most appropriate. Third party users wishing to access a restricted access datasets are required to fulfil certain criteria (e.g. be verifiably affiliated to an educational or research institution) and to sign a data access agreement which limits what can be done with the dataset. This ensures that the conditions under which the study participants consented to share their data are met.

Data deposit

In order to deposit in the University’s data repository, data.bris, the lead researcher has copied the files selected for publication into a pre-configured data-bris folder within the RDSF project space, added a readme.txt file and created a metadata record describing the deposit. Once the deposit has been checked by the Research Data Service, the metadata record has then been made public via data.bris along with an explanation of how potential users can apply for access. The constituent data are never made publicly accessible. A Digital Object Identifier (DOI) is provided so the published dataset can be easily cited.

Publication of findings

When writing up the study findings, the authors have cited the published dataset in the body of the article and provided the DOI:

Below we provide information about our reasoning when modelling the domain knowledge. We also present the resulting ontologies, data annotation and the CCBM model. The ontologies and the simulated data are publicly available¹.

¹<https://data.bris.ac.uk/data/dataset/raqa2qzai45z15b4n0za94toi>

Dataset usage

Within 2 years of publication, this dataset has received 147 unique views.

Acknowledgements

With thanks to Sam Whitehouse for insights into study data collection and selection.