Plan Overview

A Data Management Plan created using DMPonline

Title: Power BI reporting suite

Creator: Claudia Kalay

Data Manager: Claudia Kalay

Affiliation: City St George's, University of London

Template: DCC Template

Project abstract:

As part of normal reporting to Research & Enterprise Committees, reports on the number of applications and awards in departments are prepared. The data are taken from Worktribe.

The proposal is to create an automatic download of specific information from Worktribe in the form of Excel worksheets, transform them in Power BI (Desktop version) into a linked data set augmented with departmental staff FTE numbers and research income information from Finance and then publish it to the Power BI workspace in Office 365 for sharing sharing of data and more flexibility in report writing.

The downloaded data would be stored either on my personal OneDrive or on a SharePoint site. Sharepoint would make it less dependable on me. The data will not be shared outside City. Access to the database will be controlled like the access to Worktribe itself. Data sharing would be to a select group of people for whom this management information is of consideration.

The data set will include names of Worktribe users, their email address, University login, job title, affiliation to departments and research centres, and of course the bids and awards (only those within the Research & Enterprise envelope) with which they are associated. We will not download staff salary data or staff costs on projects. However, we will download information on the FTE of staff time included in projects.

ID: 111815

Start date: 21-11-2022

Last modified: 24-11-2022

Copyright information:

Power BI reporting suite

Data Collection

What data will you collect or create?

Worktribe data: applications and awards as recorded with basic details, Cost to HEI, Cost to Funder, Price to Funder, organisational affiliation, names of investigators, funder details and details of outside partners and collaborators.

Basic details will include classification for HEB-CI, HESA category, dates of bid submission and award, project tags as set up on Worktribe. We have about 5000 records on Worktribe at present and generate 600-700 per year.

We will also download user list with names, user account login and organisational affiliation. We have about 600 users on Worktribe. Only current users will be included.

We will not include any archived or discarded data.

Staff FTE data will be collated in one Excel file.
Research income data will be in Excel sheets, about 2-3 times a year.

All data will be in the format of Excel tables. In the process of data transformation, a dataset will be produced and published to Power BI Office 365 work space. Management reports will be created from this dataset.

How will the data be collected or created?

We use data downloads from Worktribe, Staff FTE information from HR system as supplied by HR and research income data as supplied by Finance. None of these data will be collected specifically for this purpose.

Searches in Worktribe will be set up to select the right data for download.

The data will be downloaded via the Worktribe API and placed into a dedicated OneDrive folder or SharePoint site.

The data will be transformed in the following ways:

- Existing column headings may be changed to something more readily understandable in the reporting context.
- A more suitable placeholder may be entered in empty fields to avoid such fields showing as 'blank' in the reports.
- Fields containing multiple information will be split up, e.g. a field containing all co-investigator names will be split up to show individual names separately
- Year date fields in the format yy/yy+1 will be introduced to mark financial years
- New tables will be created to summarise all investigators and partnerships
- Where necessary new columns will be added to imported tables to facilitate data linkage, e.g. where related data present in different formats

Data will be linked through the following identifiers:

- Department name

- School Short name
- Project ID
- Internal order number/ Finance Project code.

Neither data linkage nor data transformation will introduce new personal data.

The data will be imported into Power BI (desktop version) by the process owner where data transformation and linkage will take place. The final dataset will be published to the online Power BI workspace (either of the process owner or a team space). At the online Power BI workspace reports will be created which are connected to that dataset.

These reports may be further distributed through University work email, either as pdf or Excel tables.

Documentation and Metadata

What documentation and metadata will accompany the data?

Data are presented as is. Reports will include information on the interpretation of data. The data set will have a description outlining the content, the origin, and the regularity of update. A confidentiality note will be included.

Ethics and Legal Compliance

How will you manage any ethical issues?

Not applicable. This is management reporting and not research.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Not applicable. This is management information belonging to City and not meant to be shared outside City.

Storage and Backup

How will the data be stored and backed up during the research?

We will use City OneDrive and City SharePoint for sharing as well as workspace on Power BI in Office 365. All technical storage facilities will be City-approved storage facilities. We will follow City guidelines on use of City facilities.

How will you manage access and security?

We will use in-built features for data access, that is the normal route in Office365 to request and grant/decline data request.

Permission will be granted on a need-to-know basis depending on the role the requestor performs within the organisation.

We will make an annual check on continuing need for access.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

The dataset itself is a long-term project. It is meant to be a permanent feature to enable quicker reporting.

Data will be refreshed regularly and the data downloads used to create the dataset will be selective. Data no longer meeting the selection criteria of the downloads will be overwritten in the dataset upon refresh. At any refresh point, the dataset will conform with the selection criteria of the downloads. Any data overwritten and no longer in the selection will not be kept.

Reports drawn from the data set and located on the Power BI workspace will refresh automatically.

Any reports prepared for submission of University-mandated external reporting, e.g. HEB-CI, CASE-Ross data survey, REF, will also be stored on a SharePoint site as extracted report to preserve the data submission for archival purposes.

What is the long-term preservation plan for the dataset?

The dataset structure itself is a long-term project which will allow reports to be created easily. It is meant to be a permanent feature.

Data will be refreshed regularly and the data downloads will be selective. Data no longer meeting the selection criteria of the downloads will be overwritten. This will also refresh any reports created in the Power BI workspace.

Data Sharing

How will you share the data?

Data sharing of the final dataset and ancillary reports will happen in two ways:

Through the online Power BI workspace - an extended licence for Power BI will be necessary

By distribution of reports in pdf-format and Excel tables via University email to the report requestors.

Are any restrictions on data sharing required?

Data will only be shared on a need-to-know basis. Staff requesting data will need to make a business case why the data is needed. In most cases, this will be obvious from the role performed by staff.

Responsibilities and Resources

Who will be responsible for data management?

The process owner - currently Claudia R Kalay in her role as Head of Research Support Services. A newly created role of Research Support Services Manager (Awards, Applications & Systems) will become deputy.

What resources will you require to deliver your plan?

We rely on existing technical facilities: OneDrive, SharePoint, University email and Office 365, Worktribe.
We will need staff time of a programmer to configure the automatic download via the API.
We will need an extended licence of Power BI for at least two people.
We will need to set up a new SharePoint site.