Open Data

Contents

Overview

This document describes the Macc policy on open data.

It includes our definition of open data.

It describes our best practice for how we create and publish datasets that are available for open reuse. It details the several steps we take to ensure this best practice.

The policy also includes our Code of Conduct for those that wish to use our data.

Additionally, this document also describes how we utilise data that is already open and available.

Intended audiences
This document is for anyone interested in how Macc produces, publishes and uses open data.

Document history
This is the first issue of this policy. It was first published in October 2014.

Feedback on this policy
Feedback should be provided to Macc via email - data@macc.org.uk

Document owners
This policy is owned by the Chief Executive of Macc

Other Macc information policies
- Macc is registered with the Information Commissioner
- Macc's open data policy sits alongside policies around Confidentiality, Equality and Diversity, Social Responsibility, Website Terms of Use and staff terms and conditions.

Reuse of this policy
This policy is published under Attribution 4.0 International (CC BY-SA 4.0)
https://creativecommons.org/licenses/by/4.0/
This policy is also available for reuse via our GitHub repository:
https://github.com/MaccData/OpenDataPolicy/

About Macc
Macc is the working name of Manchester Alliance for Community Care, a charity registered in England and Wales no. 1145921
(http://apps.charitycommission.gov.uk/Showcharity/RegisterOfCharities/CharityWithPartB.aspx?RegisteredCharityNumber=1145921&SubsidiaryNumber=0) and a company limited by guarantee no. 07788593 (https://opencorporates.com/companies/gb/07788593)
Macc is registered with the Information Commissioner: Z9595019
Registered office Swan Buildings, 20 Swan Street, Manchester M4 5JW

What is open data?
Macc subscribes following definition of open data:

> Open data are data that can be freely used, reused and redistributed by anyone.
> It is subject only, at most, to the requirement to attribute and sharealike.
> Open data are the building blocks of open knowledge.
> Open knowledge is what open data becomes when it's useful, usable and used.

Macc is committed to a strong community and voluntary sector in Manchester. We believe that open data can assist with this ambition, through enabling transparency of our actions, in the form of open data.

Our open data goals
When publishing data about our work, Macc has three main aims:

That our data are useful
We wish to publish data that bring value to those building a stronger community and voluntary sector.

That our data are meaningful
We wish to publish data that add to the ongoing dialogue and discussion about the sector.

That our data are accessible
We wish to publish data in an open and accessible way, and provide regular narrative and understanding to the stakeholders.

In this context, Macc aims to operate an open data policy that is robust and practical, leading to greater engagement in the issues faced by the community and voluntary sector in Manchester.

Our open data commitments
When considering open data at Macc, we commit to the following:

We will publish data openly
- We will publish data on key aspects of our activities in an open and accessible manner
- We will update this data regularly and often
- We will provide documentation and descriptions of the data we publish, including a licence

We will use open data
- We will analyse, utilise and discuss our own data
- We will utilise other openly available datasets to provide insight into the voluntary sector of Manchester

We will support those using our data
When making our data openly available, we maintain a set of expectations, known as our Code of Conduct. If you utilise open data published by Macc, we request that you consider this.

We will support the sector with open data
In our work with voluntary sector organisations across the City of Manchester, we will provide advice and guidance in terms of publishing and using open data.

This policy provides information relevant to each of these commitments. We describe the key aspects, actions and mechanisms that we use to deliver our open data policy.

Publishing open data at Macc

Our principles and best practices
The act of providing open data is to publish and share. Macc understands that this involves responsibility and due diligence.

When Macc publishes data openly, our aim is to ensure it is of a quality to be accessed, used and understood. In doing so, we place the following expectations on our data publication.

It can often be the case that whilst data is made available openly, it is poorly structure, out-of-date, bad quality amongst other aspects. It is the intention of this policy to mitigate against such factors, via these criteria.

Note: these best practices should be read alongside our open data checklists.

Principle:	Macc open data will **respect privacy**.
Best practice:	- We will always ensure our open data is free from identifiers that could be linked to an individual person or organisation - We will not disclose any data or information that has been collected privately
See also:	Macc data privacy (Appendix 1)

Principle:	Macc open data will be **comprehensive** for the subject.
Best practice:	- We will always quality assure our data, in terms of the level of completeness and readiness for publication - We will not knowingly publish data that is incomplete for the relevant focus and/or time period
See also:	Macc data quality criteria (Appendix 1)

Principle:	Macc open data will be **relevant and succinct** for the subject.
Best practice:	- We will always consider the size, scope and spread of our data - to make it useful for those who may want to access it - We will not publish open data that is overly large, or not provide logics, lookups or additional materials
See also:	Macc data segmentation (Appendix 1)

Principle:	Macc open data will be **interoperable**.
Best practice:	- We will always provide codes and lookups for our data, particularly in terms of administrative geography - We will not publish data that involves jargon or acronyms that are not documented

See also:	Data codes checklist (Appendix 1)

Principle:	Macc open data will be presented in an **open and standard format**.
Best practice:	• We will publish data in common, accessible and standard formats such as CSV, XML • We will not publish open data in bespoke, redundant or proprietary formats
See also:	Data formats checklist (Appendix 1)

Principle:	Macc open data will be **appropriately licenced**.
Best practice:	• We will always issue an open licence with our open datasets. Our default is a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) • We will not publish data that is subject to a restrictive licence
See also:	Macc open licence (Appendix 1)

Principle:	Macc open data will always be **well documented**.
Best practice:	• We will always provide notes and guidance to accompany our datasets • We will always detail the nature, scope and purpose of our open datasets in a release document, available to all • We will not purposefully provide data that is poorly described, or requires sector knowledge to comprehend
See also:	Macc data release table (Appendix 2)

Principle:	Macc will publish data in open, **accessible** and **consistent** ways.
Best practice:	• We will always publish data in a consistent method, making it accessible to all • We will not publish data with passwords or access restrictions, or in places that are not signposted
See also:	Macc open data publishing channels (Appendix 1)

Principle:	Macc open data will be **timely**.
Best practice:	• We will always provide regular and timely updates to relevant open datasets • We will not miss updates to our relevant datasets, or let our data go "stale"
See also:	Macc data publication schedule (Appendix 1)

Principle:	Macc welcomes **feedback and discussion** of its open data.
Best practice:	• We will always make it clear how to provide feedback on our open data, and any resultant actions • We will not publish data without a feedback mechanism
See also:	Macc open data feedback mechanisms (Appendix 1)

Using open data at Macc

Alongside publishing our own datasets, Macc wishes to take an active role in the analysis and discussion of the insights that can be gleaned.

For this reason, we will establish the following as part of our open data policy.

Macc Data Blog
Alongside the formal notes and documentation on our datasets, we will maintain a blog on the main Manchester Community Central website, providing narrative and information on data being released and updated. Updates will be flagged up on our Twitter accounts and in ebulletins.

Macc staff will use these data in the form of analytical / visualisation of Macc datasets to provide insight and stimulate discussion. This will be used in the course of our work: on the Data Blog, in other staff blogs, newsletters and reports.

In the provision of our blogs and commentary, we will be open to comments and discussion, which will be published alongside any posts. These will always be moderated, in accordance with our Code of Conduct.

See also: Macc open data feedback mechanisms (Appendix 1)

Using other datasets
When we utilise other datasets in our commentary and analysis, we will always provide clear attribution and guidance as to the source of the data, and any actions we may have taken.

See also: Macc attribution checklist (Appendix 1)

Data use Code of Conduct

Macc encourages others to access, use and discuss our open data. We strive towards a strong community and voluntary sector in the City of Manchester, and value the contributions and insights that can be gleaned through use of data.

When doing so, we would hope the following basic Code of Conduct is observed:

- **Accessing our data**
  Much of the datasets published by Macc are succinct and easily available for download. When accessing our data, we request that you do not place unnecessary burden on our servers by making repeated data requests over a short period of time.

- **Attribution**
  When using our data, we request that our licence is observed. When producing any material that uses our data, please ensure an attribution to Macc is included.

- **Derivation**
  When making use of our data, always state any steps that were made to undertake calculations or analysis that are not present in the source.

- **Violations**
  When using Macc data, you must **not**:
  - make an application that pretends to be from Macc organisation;
  - present the data in a misleading or incorrect manner or to misrepresent or change the data;
  - use the name Macc or the Manchester Community Central website for party political purposes;
  - use the data in or to support a criminal or illicit activity;
  - use the data on an application to inflame or make comments that are racist, sexist or homophobic, or which promote or incite violence or illegal activity.

- **Discussion**
  We encourage discussion of our data, and the uses. In doing so, particularly in our online forums, we request you are respectful of others.

- **Feedback**
  If you spot any mistakes, errors or points for clarification, please feedback via our designated channels. We also encourage requests and ideas for new data that Macc may publish. Again, please do so via our feedback channels.

Appendix 1: Macc open data workflows

The following checklists are used by Macc staff in the preparation, publication and update of open data. These are linked to our open data best practices, detailed in our open data policy. Over time, Macc will update and enhance these.

The results of all these process will always be detailed in the data release table for the relevant dataset. An example of this is listed in Appendix 2.

When preparing any data for publication, Macc would always undertake the following:

Consideration: **Privacy**	*Action*
Does the data contain names of individuals?	If yes, then remove
Does the data contain any unique identifiers that can be used to retrieve personal information from external systems?	If yes, then remove

Consideration: **Quality**	*Action*
For data that is aggregated, check the results for any outliers	Check source data / reports to validate figures.
For any dataset, check for any blank fields or zero results.	If numerous, check source data If acceptable, then record in data release table

Consideration: **Segmentation**	*Action*
For any dataset, consider the overall physical file size	If over 10MB, then check contents and consider further segmentation
For any dataset, check the column headers and data labels are legible.	If not, provide lookup file and note in data release table
For aggregated datasets, check that aggregations are explained and logged.	Ensure these are documented in data release table
For any dataset, check that time periods used are in accordance with common standards (eg: financial quarters, calendar months)	If there is a bespoke date range, then detail in data release table

Consideration: **Data Codes**	*Action*
For geographic areas used within datasets, provide the code alongside the	Applicable to:

name.	Local authority Ward Clinical Commissioning Groups
Provide and/or signpost data users to the latest lookup of any codes used.	In the case of administrative geographic regions, refer to authoritative sources such as Ordnance Survey, Office for National Statistics and the NHS
When using internal / Macc specific codes, ensure that a lookup and/or explanation is provided.	Log this in the data release table

Consideration: **Formats**	*Action*
For spreadsheets and tabular data, release in standard open formats.	Release as: Open Document Format for spreadsheets (.odf) Comma Separated Format for flat files (.csv)`
When working with other data standards and systems, ensure that the format is open and accessible.	Consider XML, JSON or RDF formats as open. Check with standard or publication organisation
Avoid publishing data in closed, proprietary and formats that make the data inaccessible.	

When publishing data, Macc will always ensure a relevant licence is provided.
Our default licence is a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Consideration: **licence**	*Action*
Does the dataset fit within the default licence for Macc? See: http://creativecommons.org/licenses/by-sa/4.0/	If yes, then ensure this is in the data release table If no, then select alternative, and document accordingly
Does the data contain any information that is derived from other sources?	If yes, then detail these sources in the data release table If there may be an issue with these derivations, then seek advice

Consideration: **Schedule**	Action
For datasets that are updated periodically, ensure this takes place within acceptable timeframe. Guidance: Monthly updates - two weeks after period ends Quarterly updates - one month after period ends Annual updates - three months after period ends	If data publication is outside of these thresholds, update/add to data release table
Ensure that relevant older data can be accessed after an update - that it is not deleted or destroyed.	For ongoing statistics, ensure the new time period data is made available alongside other periods Where data must be overwritten, document in data release table

Consideration: **Publishing Channels**	Action
When creating data files, check the file name is logical, descriptive and devoid of any inconsistencies. eg: Macc_volunteers_statistics.csv rather than Macc_volunteersstats(1)-EDITED.csv	If a file naming convention is to be created, then update accordingly
When hosting data files, always ensure that the end URL is accessible, and free from any security barriers, passwords or blocks.	If there is an issue in terms of accessing the URL to the data file, seek advice

Consideration: **Feedback**	Action
When an "Issue" is posted via online forums (such as GitHub), respond accordingly - even if acknowledgement.	Where the issue can be progressed, respond accordingly If no action can be applied immediately, respond accordingly
When feedback is received via settings such as face-to-face meetings or	Where appropriate, create a new Issue for the relevant dataset, attributing the

workshops, consider how best to add this to existing feedback.	source of the observation / remark
When comments are made about Macc usage of open data, respond according to the Code of Conduct.	

When using datasets published by other organisations, Macc will always ensure the following attribution considerations are provided within the context of any material we produce.

Attribution	*Comment*
Name of the dataset utilised	e.g.: Adult Learning Centres
Publisher of dataset	e.g.: Manchester City Council
Source URL (from where the data can be retrieved)	e.g.: http://www.manchester.gov.uk/site/custom_scripts/getmetadata.php?guid=a65c8dfb-7371-4db9-a3b3-a9e4be72a493
Date retrieved	DD/MM/YYYY
Notes on usage	Any notes on actions undertaken that result in the source data being changed or modified

Appendix 2: data release table (template)

Example: Volunteer data

Data consideration	Macc response
Dataset name	Volunteer data
Brief description	Quarterly statistics on the volunteers registered with the Volunteer Centre
Responsible person / team *Who is the authority for this data?*	Volunteer team
Privacy *What privacy considerations should be taken?*	Aggregate data only. Names and personal details are not included. Unique identifiers are omitted Data is not cross tabulated
Legality *Does the data contain any derived data?*	No
Format How will be the data be made available	As a spreadsheet, saved in a CSV format
Segmentation *How will the data be segmented?*	One file, with the following reports: • Count of new volunteers, per quarter • Count of new volunteers, by ward / area, per quarter • Count of new volunteers, by gender, per quarter
Dates *What date period are used in the data, if any?*	The data will be presented in financial year quarters: • FY Q1 - 1 April - 30 June • FY Q2 - 1 July - 30 September • FY Q3 - 1 August - 31 December • FY Q4 - 1 January - 31 March
Structured *Does the data contain any inherent jargon or interpretation?*	No. 'New volunteers' will be those that registered with the Volunteer Centre during the above periods
Interoperable *How can other datasets be referenced?*	Ward codes are included as both nine - in both ONS and GSS formats

Availability *How will the data be disseminated?*	The Macc open data website Registered on DataGM
Documentation *How will the data (and processes) be described?*	A blog post on Macc data website Release notes on DataGM
Updated *What is the update timetable?*	Quarterly - two weeks after the end of each financial year quarter
Licencing *What licence will be applied?*	Creative Commons, Attribution http://creativecommons.org/licenses/by/4.0/
Narrative *What is planned in terms of usage and blogs?*	Outreach team to post three blog posts...
Feedback *How will feedback be progressed?*	Via outreach sessions Blog comments GitHub issues (technical community) Usual Macc feedback processes
Additional Notes	The postcode provided by volunteers is used to determine the ward / area. This is recorded on the volunteer database only, and not contained in the published open data

Appendix 3: Macc open data sets

List of standard datasets released by Macc. This list will be updated and maintained.

Dataset	Description	First published	Update(s)
Volunteer data	Aggregate statistics on volunteers		Quarterly
Groups data	Aggregate statistics on groups registered with Macc		Quarterly
Funding data	Details of funding provided by Macc to other organisations		Quarterly
State of the sector survey	Anonymised responses to annual survey		Annually

Other datasets may be released from time to time as they become available and if they are of interest.

Policy Checklist
Conflict of Interest Policy

Date first adopted: 24ᵗʰ October 2014

Review Dates:

Date of review	Amendments/Updates made	Reviewed & accepted as is	Proposed next review date
24/10/2014	Policy signed off by Chief Executive.		October 2015