Best Practices in Data Analysis and Sharing in Neuroimaging using MRI

Appendix 2.

This section contains checklists for practices and items to report. Each item has been included because it is an essential piece of information needed to understand, evaluate and reproduce an experiment. Authors should strive to include all these items, but items marked as "Mandatory" are particularly crucial, and a published work cannot be considered complete without such information.

Experimental Design Reporting

Aspect	**Notes**	**Mandatory**
Number of subjects	*Elaborate each by group if have more than one group.*	
Subjects approached		N
Subjects consented		N
Subjects refused to participate	Provide reasons.	N
Subjects excluded	Subjects excluded after consenting but before data acquisition; provide reasons.	N
Subjects participated and analyzed	Provide the number of subjects scanned, number excluded after acquisition, and the number included in the data analysis. If they differ, note the number of subjects in each particular analysis.	Y
Inclusion Criteria and Descriptive Statistics	*Elaborate each by group if have more than one group.*	
Age	Mean, standard deviation and range.	Y
Gender	Absolute counts or relative frequencies	Y

Race & Ethnicity	Per guidelines of NIH or other relevant agency	N
SES, Education	Specify measurement instrument used; may be parental SES and education if study has minors.	N
IQ	Specify measurement instrument used.	N
Handedness	Absolute or relative frequencies; basis of handedness-attribution (self-report, EHI, other tests)	Y
Exclusion criteria	Describe any screening criteria, including those applied to “normal” sample such as MRI exclusion criteria.	Y
Clinical criteria	Detail the area of recruitment (in- vs. outpatient setting, community hospital vs. tertiary referral center etc.) as well as whether patients were currently in treatment.	Y
Clinical Instruments	Describe the instruments used to obtain the diagnosis and provide tests of intra- or inter-rater reliability. Clarify whether a “clinical diagnosis” or “inventory diagnosis” was used (if applicable). State the diagnostic system (ICD, DSM etc) that was used.	
Matching strategy	If applicable.	Y
Population & Recruitment strategy	Population from which subjects were drawn, and how and where recruitment took place, e.g., schools, clinics, etc. If possible, note if subjects are research-naive have participated in other studies before.	Y
Subject scanning order	With multiple groups, use of randomized or interleaved order to avoid bias due to scanner changes/upgrades	N
Neurocognitive measures	All measures collected on subjects should be described and reported.	Y
Ethical Considerations		

Ethical Approval	Describe approval given, including the particular institutional review board, medical ethics committee or equivalent that granted the approval. When data is shared, describe the ethics/institutional approvals required from either the author (source) or recipient.	Y
Informed consent	Record whether subjects provided informed consent or, if applicable, informed assent.	Y
Design Specifications		
Design type	Task or resting state. Event-related or block design. (See body text for usage of 'block design' terminology.)	Y
Condition & Stimuli	Clearly describe each condition and the stimuli used. Be sure to completely describe baseline (e.g. blank white/black screen, presence of fixation cross, or any other text), especially for resting-state studies. When possible provide images or screen snapshots of the stimuli.	Y
Number of blocks, trials or experimental units	Specify per session, and if differing by subject, summary statistics (mean, range and/or standard deviation) of such counts.	Y
Timing and duration	Length of each trial or block (both, if trials are blocked), and interval between trials. Provide the timing structure of the events in the task, whether a random/jittered pattern or a regular arrangement; any jittering of block onsets.	Y
Length of the experiment	Describe the total length of the scanning session, as well as the duration of each run. (Important to assess subject fatigue.)	Y
Design optimization	Whether design was optimized for efficiency, and how.	N
Presentation software	Name software, version and operating system on which the stimulus presentation was run. When possible, provide code (see replication & reproducibility section).	Y
Task specification		

Condition	Enumerate the conditions and fully describe and reference each. Consider using a shorthand name, e.g. AUDSTIM, VISSTIM, to refer to each condition, to clarify the distinction between a specific modeled effect and a psychological construct. Naming should reflect the distinction between instruction periods and actual stimuli, and between single parameters and contrasts of parameters.	Y
Instructions	Specify the instructions given to subjects for each condition. For resting-state, be sure to indicate eyes-closed, eyes-open, any fixation. Describe if the subjects received any rewards during the task, and state if there was a familiarization / training inside or outside the scanner.	Y
Stimuli	Specifics of stimuli used in each run. For example, the unique number of stimuli used, and whether/how stimuli were repeated over trials or conditions.	Y
Randomization	Describe block or event ordering as deterministic, or report manner of randomization, in terms of order and timing. If pseudo-randomized, i.e. under constraints, describe how and the criteria used to constrain the orders/timings.	Y
Behavioral Performance		
Variables recorded	State number of type of variables recorded (e.g. correct button press, response time).	Y
Summary statistics	Summaries of behavior sufficient to establish that subjects were performing the task as expected. For example, correct response rates and/or response times, summarized over subjects (e.g. mean, range and/or standard deviation).	Y

2. Acquisition Reporting

Aspect	**Notes**	**Mandatory**
Subject preparation		

Mock scanning	Use of a MRI simulator to acclimate subjects to scanner environment. Report type of mock scanner and protocol (i.e. duration, types of simulated scans, experiments).	N
Special accommodations	For example, for pediatric scanning, presence of parent/guardian in the room.	Y
MRI system description		
Scanner	Provide make, model & field strength in Tesla.	Y
Coil	Receive coil (e.g. “a 12-channel phased array coil”, but more details for a custom coil) and (if nonstandard) transmit coil. Additional information on the gradient system, e.g. gradient strength (if non-standard for the make and model, or switchable).	Y
Significant hardware modifications	For example, special gradient inserts/sets.	N
Software version	Highly recommended when sharing vendor-specific protocols or exam cards, as version may be needed to correctly interpret that information.	N
MRI Acquisition: General		
Pulse sequence type	For example, gradient echo, spin echo, etc.	Y
Imaging type	For example, echo planar imaging (EPI), spiral, 3D. Number of shots (if multi-shot); partial Fourier scheme & reconstruction method (if used);	Y
Essential sequence parameters.	For all acquisitions: ● Echo time (TE) ● Repetition time (TR) ● Flip angle (FA) Functional MRI: ● Number of volumes	Y

	• Sparse sampling delay (delay in TR) if used Inversion recovery sequences: • Inversion time (TI) B0 field maps: • Echo time difference (dTE) Diffusion MRI: • Number of directions • b-values • Number of b=0 images • Number of averages (if any) • Single- or dual-spin-echo, gradient mode (serial or parallel)	
Phase encoding direction	Mention if phase encoding reversal (“blip-up/blip-down”) is used. For 3D, specify “partition encode” (aka slice) direction.	Y
Parallel imaging method & parameters	E.g. SENSE, GRAPPA or other parallel imaging method, and acceleration factor. Matrix coil mode, and coil combining method (if non-standard; Siemens-specific).	Y
Multiband parameters	Multiband factor and field-of-view shift (only if applicable).	Y
Readout bandwidth		N
Fat suppression	Especially for anatomical scans, whether it was used or not.	N
Essential imaging parameters.	Field of view; for 2D acquisitions, in-plane matrix size, slice thickness and interslice gap; for 3D acquisitions, 3D matrix size.	Y
Slice order & timing	For fMRI acquisitions, interleaved vs. sequential ordering and direction (ascending/descending), location of 1st slice; any specialized slice timing.	N
Acquisition orientation	Axial, sagittal, coronal, or oblique. If axial and co-planar w/ AC-PC, the volume coverage in terms of Z in mm.	Y
Slice position procedure	For example, landmark guided vs. auto-alignment.	N

Brain coverage	Report whether coverage was whole-brain, and whether cerebellum and brainstem were included. If not whole-brain, note the nature of the partial area of coverage.	Y
Scanner-side preprocessing	Examples include prospective-motion correction, signal inhomogeneity correction, distortion-correction, or a reconstruction matrix size that differs from acquisition matrix.	Y
Scan duration		N
Other non-standard procedures	For example, turning the cold head(s) off during EEG/fMRI or spectroscopy measurements, or reduce sound pressure by limiting the gradient slew rate.	N
T1 stabilization	Number of initial “dummy” scans acquired and then discarded by the scanner.	Y
Diffusion MRI gradient table	Also referred to as the b-matrix.	N
Preliminary Quality Control		
Motion monitoring	For functional or diffusion acquisitions, any visual or quantitative checks for severe motion; likewise, for structural images, checks on motion or general image quality.	N
Incidental findings	Protocol for review of any incidental findings, and how they are handled in particular with respect to possible exclusion of a subject’s data.	N

3. Preprocessing Reporting

Aspect	**Notes**	**Mandatory**
Software	For each software used, be sure to include version and revision number.	Y
Software citation	Include URL and Research Resource Identifier for each software used.	N

T1 stabilization	Number of initial "dummy" scans discarded as part of preprocessing (if not already performed by scanner).	Y
Brain extraction	If performed, report: ● Name of software/method (e.g., BET, recon-all in FreeSurfer, etc) ● Parameter choices (e.g. BET's fractional intensity threshold) ● Any manual editing applied to the brain masks	Y
Segmentation	For structural images, method used to extract gray, white, CSF and other tissue classes.	Y
Slice time correction	If performed, report: ● Name of software/method ● Whether performed after or before motion correction ● Reference slice ● Interpolation type (e.g., spline or sinc)	Y
Motion correction	Report: ● Name of software/method ● Use of non-rigid registration, and if so the type of transformation ● Use of motion susceptibility correction (fieldmap-based unwarping), as well as the particular software/method ● Reference scan (e.g. 1st scan or middle scan) ● Image similarity metric (e.g. normalized correlation, mutual information, etc) ● Interpolation type (e.g., spline, sinc) ● Use of any slice-to-volume registration methods, or integrated with slice time correction	Y
Function-Structure (intra-subject) coregistration	Report: ● Name of software/method ● Type of transformation (rigid, nonlinear); if nonlinear, type of transformation ● Cost function (e.g., correlation ratio, mutual information, boundary-based registration, etc) ● Interpolation method (e.g., spline, linear)	Y

	Note this step might not be necessary if direct T2* to a functional template registration is used.	
Distortion correction	Use of any distortion correction due to field or gradient nonlinearity.	Y
Intersubject registration	Report: ● Name of software/method (e.g., FSL flirt followed by fnirt) ● Whether volume and/or surface based registration is used (if not already clearly implied) ● Image types registered (e.g. T2* or T1) ● Any preprocessing to images; e.g. for T1, bias field correction, or segmentation of gray matter; for T2*, single image (specify image) or mean image ● Template space (e.g., MNI, Talairach, fsaverage, FS_LR), modality (e.g., T1, T2*), resolution (e.g., 2mm, fsaverage5, 32k_FS_LR), and the specific name of template image used; note the domain of the template if not whole brain, i.e. cortical surface only, cerebellum only, CIFTI 'grayordinates' (cortical surface vertices + subcortical gray matter voxels), etc. ● Additional template transformation for reporting; e.g., if using a template in MNI space, but reporting coordinates in Talairach, clearly note and report method used (e.g., Brett's mni2tal, Lancaster's icbm_spm2tal) ● Choice of warp (rigid, nonlinear); if nonlinear, transformation type (e.g., B-splines, stationary velocity field, momentum, non-parametric displacement field); if a parametric transformation is used, report resolution, e.g., 10x10x10 spline control points ● Use of regularization, and the parameter(s) used to set degree of regularization ● Interpolation type (e.g., spline, linear); if projection from volume to surface space, how were voxels sampled from the volume (e.g., trilinear; nearest neighbor; ribbon-constrained specifying inner and outer surface used) ● Cost function (e.g., correlation ratio, mutual information, SSD) ● Use of cost-function masking	Y

Intensity correction	Bias field corrections for structural MRI, but also correction of odd versus even slice intensity differences attributable to interleaved EPI acquisition without gaps.	Y
Intensity normalization	Scan-by-scan or run-wide scaling of image intensities before statistical modelling. E.g. SPM scales each run such that the mean image will have mean intracerebral intensity of 100; FSL scales each run such that the mean image will have an intracerebral mode of 10,000.	N
Artifact and structured noise removal	Use of physiological noise correction method. Report: ● Name of software/method used (e.g. CompCor, ICA-FIX, etc) ● If using a nuisance regression method, specify regressors used; for each type, include key details, as follows: ○ Motion parameters ■ Expansion basis and order (e.g. 1st temporal derivatives; Volterra kernel expansion) ○ Tissue signals ■ Tissue type (e.g., whole brain, gray matter, white matter, ventricles) ■ Tissue definition (e.g., a priori seed, automatic segmentation, spatial regression) ■ Signal definition (e.g., mean of voxels, first singular vector, etc.) ○ Physiological signals ■ e.g., heart rate variability, respiration	Y
Volume censoring	Remediation of problem scans, also known as “scrubbing” or “de-spiking”. Report: ● Name of software/method ● Criteria (e.g., frame-by-frame displacement threshold, percentage BOLD change) ● Use of censoring or interpolation; if interpolation, method used (e.g., spline, spectral estimation)	Y

Resting state fMRI feature	Creation of summary measure like ALFF, fALFF, ReHo. For ALFF, fALFF report: ● Lower and upper band pass frequencies For ReHo, report: ● Neighborhood size used to compute local similarity measures (e.g. 7, 19 or 27) ● Similarity measure (e.g. Kendall's coefficient of concordance)	Y
Spatial smoothing	If this preprocessing step is performed, report: ● Name of software/method ● Size and type of smoothing kernel ● Filtering approach, e.g., fixed kernel or iterative smoothing until fixed FWHM ● Space in which smoothing is performed (i.e. native volume, native surface, MNI volume, template surface)	Y
Quality Control reports	Summaries of subject motion (e.g. mean FD), image variance (e.g. DVARS), and note of any other irregularities found (e.g. motion or poor SNR not sufficiently severe to warrant exclusion).	N

4. Statistical Modeling & Inference

Aspect	**Notes/Ontology**	**Mandatory**
Mass Univariate Analyses		
Dependent variable - data submitted to statistical modeling	Report the number of time points, number of subjects; specify exclusions of time points / subjects, if not already specified in experimental design.	Y
Dependent variable - Spatial region modeled	If not "Full brain", give a specification of an anatomically or functionally defined mask.	Y

Independent variables	For first level fMRI, specify: ● Event-related design predictors ○ Modeled duration, if other than zero. ○ Parametric modulation ● Block Design predictors ○ Note whether baseline was explicitly modeled ● HRF basis, typically one of: ○ Canonical only ○ Canonical plus temporal derivative ○ Canonical plus temporal and dispersion derivative ○ Smooth basis (e.g. SPM “informed” or Fourier basis; FSL’s FLOBS) ○ Finite Impulse Response model ● Drift regressors (e.g. DCT basis in SPM, with specified cut-off) ● Movement regressors; specify if squares and/or temporal derivative used ● Any other nuisance regressors, and whether they were entered as interactions (e.g. with a task effect in 1st level fMRI, or with group effect) ● Any orthogonalization of regressors, and set of other regressors used to orthogonalize against For second level fMRI or general group model, specify: ● Group effects (patients vs. controls) ● Clearly state whether or not covariates are split by group (i.e. fit as a group-by-covariate interaction) ● Other between subject effects (age, gender; for VBM, total GM or ICV).	Y
Model type	Some suggested terms include: ● “Mass Univariate” ● “Multivariate” (e.g. ICA on whole brain data) ● “Mass Multivariate” (e.g. MANOVA on diffusion or morphometry tensor data) ● “Local Multivariate” (e.g. “searchlight”) ● “Multivariate, intra-subject predictive” (e.g. classify individual trials in event-related fMRI)	Y

	● "Multivariate inter-subject predictive" (e.g. classify subjects as patient vs. control) ● "Representation Similarity Analysis"	
Model settings	The essential details of the model. For mass-univariate, first level fMRI, these include: ● Drift model, if not already specified as a dependent variable (e.g. locally linear detrending of data & regressors, as in FSL) ● Autocorrelation model (e.g. global approximate AR(1) in SPM; locally regularized autocorrelation function in FSL) For mass-univariate second level fMRI these include: ● Fixed effects (all subjects' data in one model) ● Random or mixed-effects model, implemented with: ○ Ordinary least squares (OLS, aka unweighted summary statistics approach; SPM default, FSL FEAT's "Simple OLS") ○ weighted least squares (i.e. FSL FEAT's "FLAME 1"), using voxel-wise estimate of between subject variance ○ Global weighted least squares (i.e. SPM's MFX) With any group level model (e.g. fMRI, VBM, etc) be sure to indicate any repeated measures structure, and how this was modeled.	Y
Inference: Contrast/effect	Specification of the precise effect tested, often as a linear contrast of parameters in a model. When possible, define these in terms of the task or stimulus conditions instead of psychological concepts (See *Task Specification* in *Experimental Design Reporting*). Include any details of masking and/or conjunction with other contrasts. Indicate any use of any omnibus ANOVA tests.	Y
Inference: Search region	Whole brain or "small volume"; carefully describe any small volume correction used for each contrast. If a small-volume correction mask is defined anatomically, provide named anatomical regions from a publicly available ROI atlas. If small-volume correction mask is functionally define, clearly describe the functional task and identify any risk of circularity.	Y

Inference: Statistic type	For mass-univariate, typically one of ● Voxel-wise (aka peak-wise in SPM) ● Cluster-wise ○ Cluster size ○ Cluster mass ○ Threshold-free Cluster Enhancement (TFCE) For cluster size or mass, be sure to provide cluster-forming threshold; likewise indicate use of non-standard TFCE parameters.	Y
Inference: P-value computation	Note if anything but standard parametric inference used to obtain P-values. If nonparametric method was used, report method (e.g. permutation or bootstrap) and number of resamples used.	
Inference: Multiple testing correction	For mass-univariate, specify the type of correction and how it is obtained, especially if not the typical usage. Usually one of ● Familywise Error ○ Random Field Theory (typical) ○ Permutation ○ Bonferroni ● False Discovery Rate ○ Benjamini & Hochberg FDR (typical) ○ Positive FDR ○ Local FDR ○ Cluster-level FDR ● None/Uncorrected	Y
Functional Connectivity		
Confound Adjustment	Method for detecting movement artifacts, movement-related variation, and remediation (e.g. ‘scrubbing’, ‘despiking’, etc). Use of global signal regression, exact type of global signal used and how it was computed.	Y

Multivariate Method, Independent Component Analysis	Report: ● Algorithm to estimate components ● Number of components (if fixed), or algorithm for estimating number of components ● If used, method to synthesize multiple runs ● Sorting method of IC's, if any ● Detailed description of how components were chosen for further analysis	Y
Dependent variable definition	For seed-based analyses report: ● Definition of the seed region(s) ● Rationale for choosing these regions For region-based analyses report: ● Number of ROIs ● How the ROI's are defined (e.g. citable anatomical atlas; auxiliary fMRI experiments); note if ROIs overlap ● Note if considering only homotopic connectivity [Stark2008] ● Assignment of signals to regions (i.e. how a time series is obtained from each region, e.g. averaging or first singular vector)	Y
Functional connectivity measure/ model	Report: ● Measure of dependence used, e.g. Pearson's (full) correlation, partial correlation, mutual information, etc; also specify ○ Use of Fisher's Z-transform (Yes/No) and, if standardised, effective N is used to compute standard error (to account for any filtering operations on the data) ○ Estimator used for partial correlation ○ Estimator used for mutual information ● Regression model used to remove confounding effects (Pearson or partial correlation)	Y
Effectivity connectivity	Report: ● Model ● Algorithm used to fit model ● If per-subject model, method used to generalize inferences to population	Y

	● Itemize models considered, and method used for model comparison	
Graph analysis	Report the 'dependent variable' and 'functional connectivity measure' used (see above). Specify either ● Weighted graph analysis, or ● Binarized graph analysis is used, clarifying the method used for thresholding (e.g. a 10% density threshold, or a statistically-defined threshold); consider the sensitivity of your findings to the particular choice of threshold used Itemise the graph summaries used (e.g. clustering coefficient, efficiency, etc), whether these are global or per-node/per-edge summaries. In particular with fMRI or EEG, clarify if measures applied to individual subject networks or group networks.	Y
Multivariate Predictive Analysis		
Independent variables	Specify the ● Variable type (discrete or continuous) ● Class proportions in classification settings If available, report on population stratification: ● Provide information on the population; specify how this is taken into account in the predictive model	Y
Features extracted from data	Report the method used for feature transformation and selection. When data-driven methods are used, specify the data handling procedures to learn the parameters involved.	Y
Learning method	Describe the method used and parameters settings, and report how the convergence of the learning method is monitored.	Y
Training procedure	Provide a description of the	Y

	● Pipeline structure applied uniformly to all cases (e.g. that could be independently applied to a new case) ● Method for hyper-parameter setting ● Data splitting (cross validation)	
Evaluation criterion	Include a precise description and justification of the evaluation criterion used. In particular, including the method used to estimate the significance of the accuracy score.	Y
Feature importance	Procedure used to estimate the importance of the features (if feature importance is investigated) in the classification engine.	Y

5. Results Reporting

Aspect	**Notes/Ontology**	**Mandatory**
Mass Univariate analysis		
Effects tested	Provide a complete list of tested and omitted effects	Y
Tables of coordinates	Provide one table of coordinates including ● Contrast / effect to which it refers ● Anatomical region, XYZ coordinate (with coordinate system, MNI, Talairach, noted in caption, and anatomical region (in caption or body text, describe source of labels, e.g. subjective, atlas, etc) ● P-value forming basis of inference (e.g. voxel-wise FWE corrected P; or cluster-wise FDR corrected P) ● T/Z/F statistic (with degrees of freedom in table caption) ● In caption, state whether coordinates are from whole brain, or from a specific constrained volume.	Y

Thresholded maps	For each effect, provide images of maps of significant regions, ensuring that each caption describes: ● Type of inference and the correction method, as well as form of any sub-volume corrections applied when computing corrected significance. ● Include color bars; when presenting multiple maps in a figure, use a common color bar to ensure the results are comparable	N
Extracted data	State whether data extracted from an ROI (e.g. to compute an effect size) is defined based on independent data, as otherwise it is susceptible to bias	Y
Spatial Features	Report the ● Size of the analysis volume in voxels, mm ● Spatial smoothness of noise (e.g. FWHM) and Resel count (if using Random Field Theory)	Y
Functional connectivity		
ICA analyses	Report the total number of components (especially when estimated from the data and not fixed). Report the number of these analyzed and the reason for their selection.	Y
Graph analyses: Null hypothesis tested	For graph-based methods, carefully state what is the null hypothesis of the test and how the statistic distribution under the null is computed.	Y
Multivariate Predictive Analysis		
Evaluation Metrics: Discrete Response	Always report: ● Accuracy ● If group sizes unequal, balanced (or average) accuracy. When there are only 2 classes, and one can be labeled “positive” ● Precision (1 − false discovery rate) ● Recall (sensitivity) ● False positive rate (1-specificity)	Y

	● F1 (incorporates both precision and recall) ● Receiver operating characteristic (ROC) curves, e.g. summarised by area under the curve (AUC); AUC for only high specificity (e.g. false positive rates no greater than 10%) are also useful When there are 3 or more classes, be sure to provide the confusion matrix.	
Evaluation Metrics: Continuous Response	"Prediction R^2", percentage of variance explained by prediction. (Note this *is not* the squared correlation coefficient between true and predicted values).	Y
Evaluation Metrics: Representational similarity analysis	Report the Kendall Tau statistic for each candidate model considered.	Y
Evaluation Metrics: Significance	When possible use formal test to obtain P-value to assess whether evaluation metric is "significant" or consistent with noise.	Y

6. Data Sharing

Aspect	**Notes**	**Mandatory**
Reporting a Data Sharing Resource		
Material shared	List types of images and non-imaging data provided. Report on the completeness of the data (e.g., number of subjects where all types of imaging, demographic, and behavioral data is available).	Y
URL, access information	● Stable URL or DOI ● Specific instructions on how to gain access. Specifically mention whether application must be vetted for particular intended research use (e.g. to preclude multiple users investigating the same question), or whether a research collaboration is must be established. ● Cost of access	Y

Ethics Compliance	Confirm that the ethics board of the host institution generating the data approves the sharing of the data made available. Clarify any constraints on uses of shared data, for example, whether users downloading the data also need ethics approval from their own institution..	Y
Documentation	Provide URL to documentation, and specify its scope (e.g. worked examples, white papers, etc).	N
Data Format	Report the format of the image data shared, e.g. DICOM, MINC, NIFTI, etc.	Y
Ontologies	Data organization structures, including Data Dictionaries and Schemas. Is the software using an established ontology?	N
Visualization	Availability of in-resource visualization of the imaging or non-imaging data.	N
De-identification	How, if at all, data are de-identified.	N
Provenance and history	Availability of detailed provenance of preprocessing and analysis of shared data.	N
Interoperability	Ability of a repository to work in a multi-database environment, availability of API's and ability to connect to analysis pipelines.	N
Querying	Mechanisms available for constructing queries on the repository (e.g. SQL, SPARQL).	N
Versioning	How users can check version of downloaded data and compare it to the current version at a later time.	N

7. Reproducibility

Aspect	**Notes/Ontology**	**Mandatory**
Documentation		

Tools used	Tool names, versions, and URLs.	Y
Infrastructure	Machine CPU model, operating system version, any use of parallelization.	Y
Workflow	Use of a work flow system, its version and URL.	N
Provenance trace	State whether detailed provenance information is available.	N
Literate program implementing results	Provide a URL linking to the relevant resource; for example, an ipython notebook implementing key analyses.	N
Archiving		
Tools availability	Note if tools are publically available.	N
Virtual appliances	Note if a virtual environment to facilitate repeat analysis is available.	N
Citation		
Data	Provide permanent identifier if possible.	N
Workflow	Provide permanent identifier if possible.	N