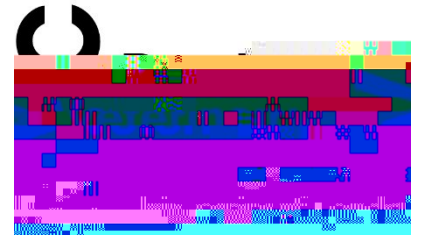
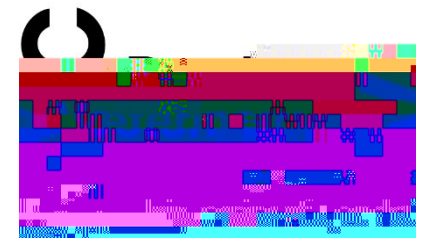


Introduction to Research Data Management



Aims of course





What is research data?

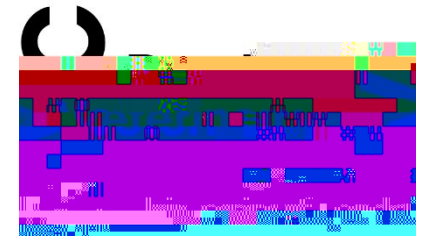
“representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship” includes:

source data - all data collected, created and used by the research, including data held elsewhere

assembled datasets - data extracted or derived from the above

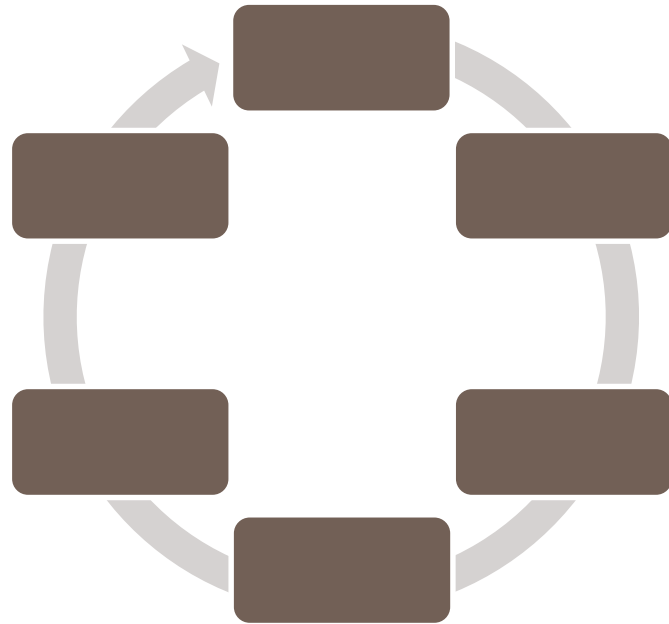
referenced data - any subset of the above that has been used in analysis or to draw conclusions. Consistent with whatever is considered ‘supplementary material’ to research findings in your domain.”

'Categories' of research data

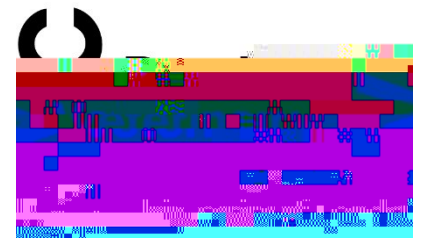
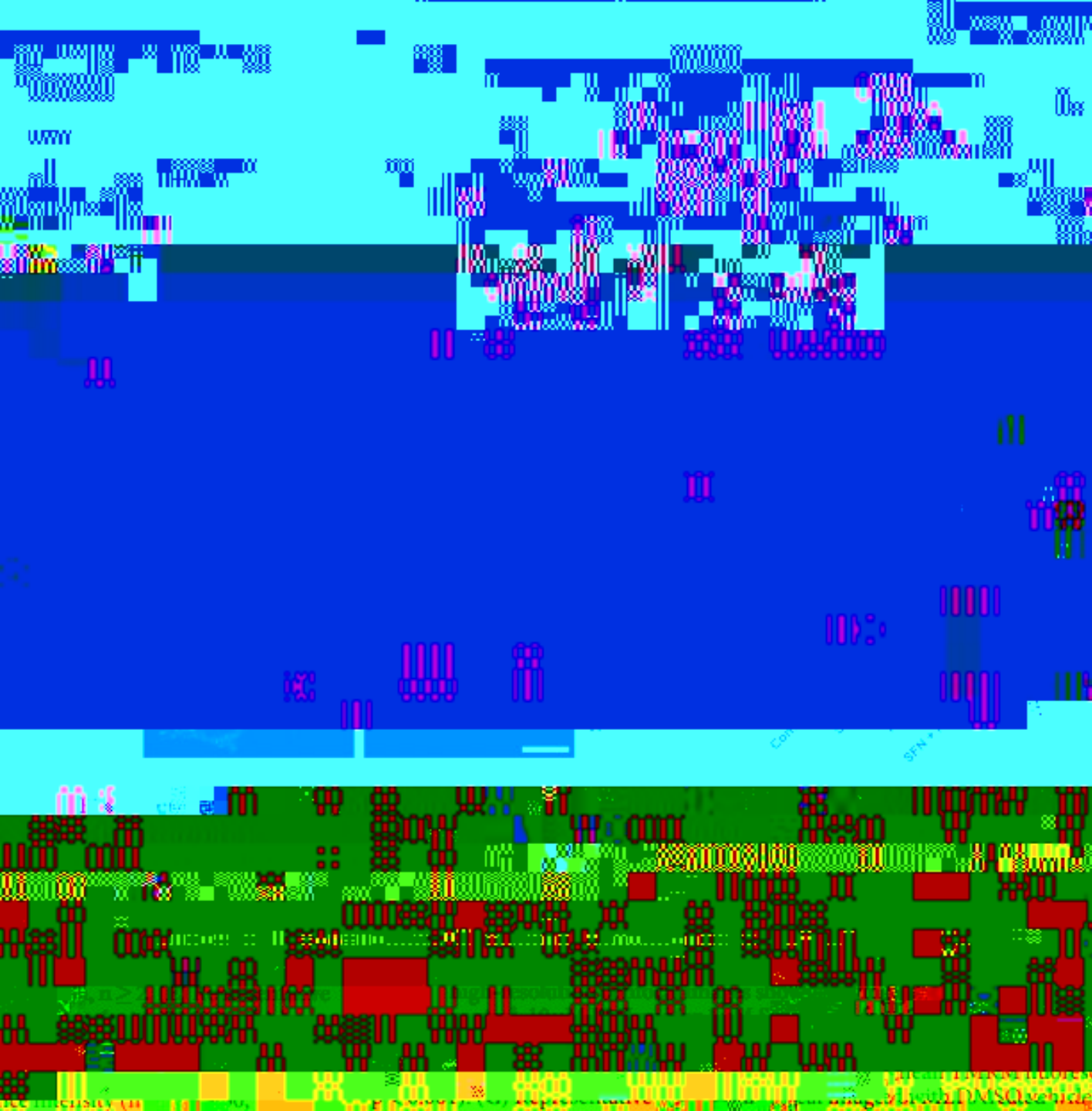


What is research data management?

The "organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results."



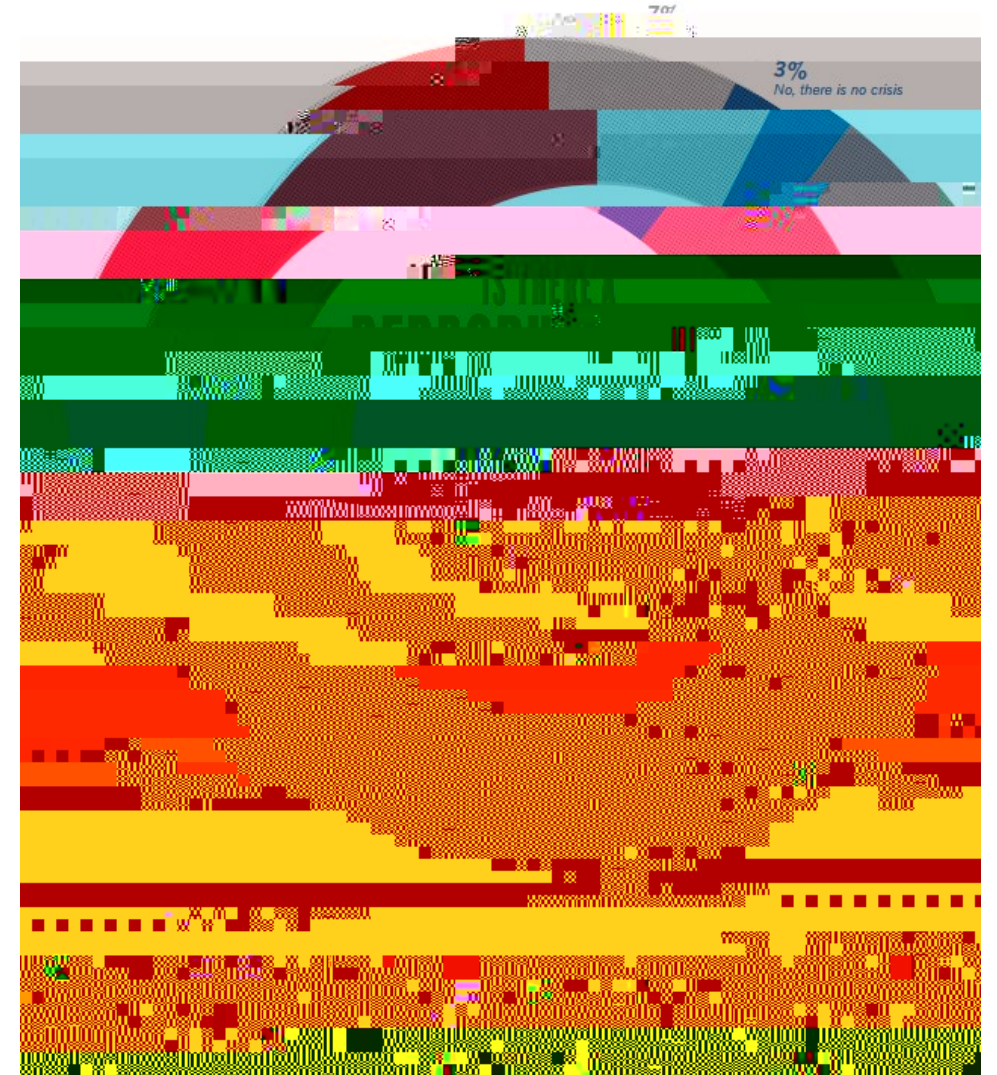
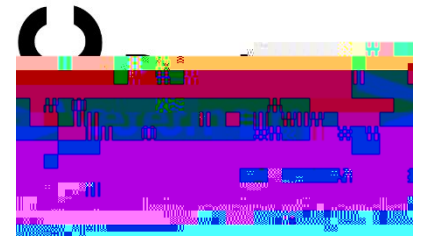
Making the case for Research Data Management

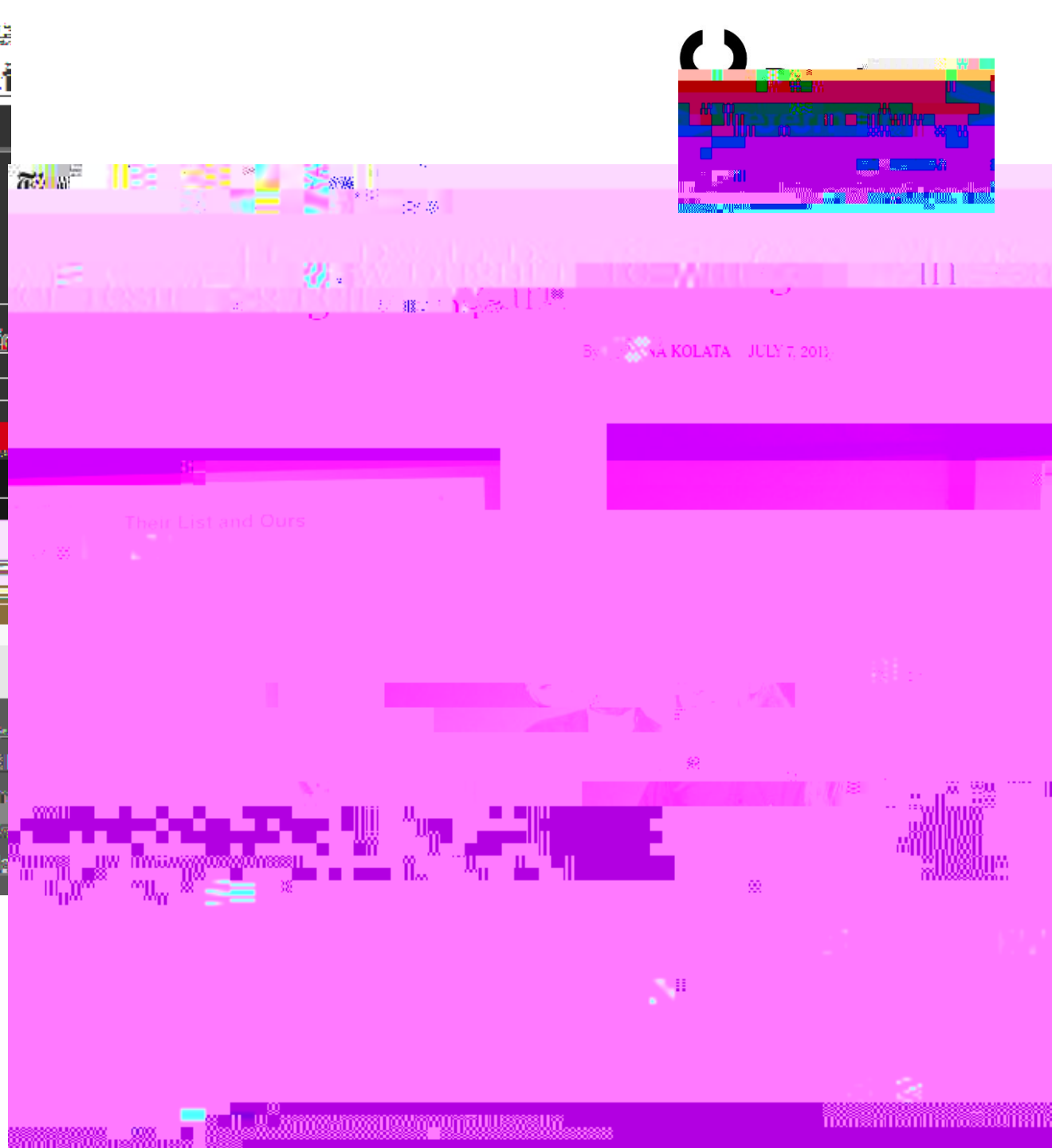
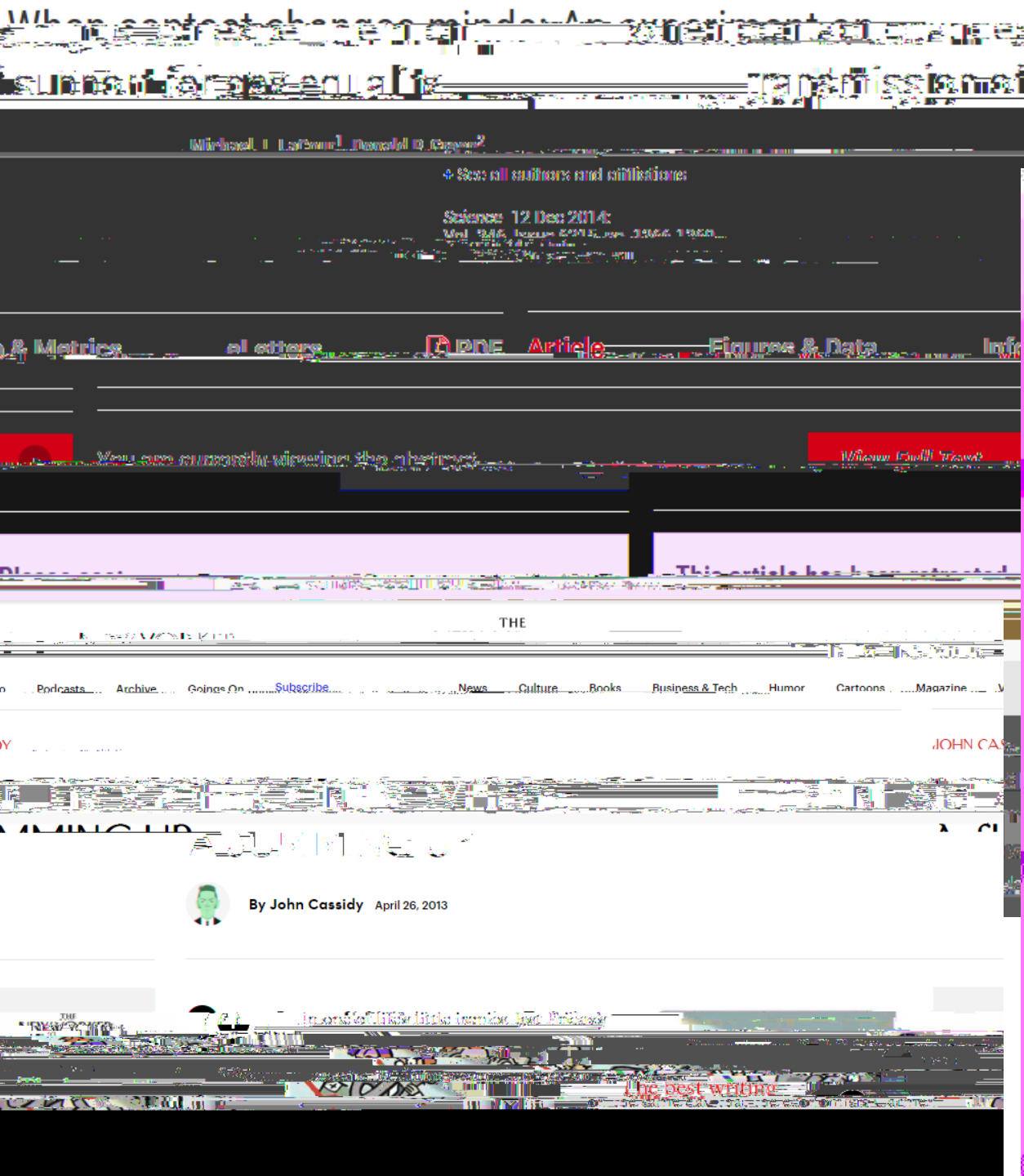


'Representation'

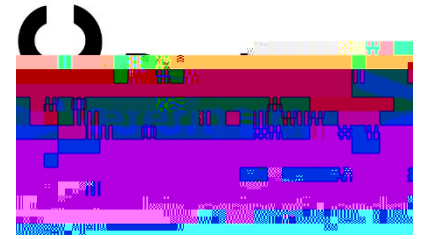
Reversible

Reproducibility of research

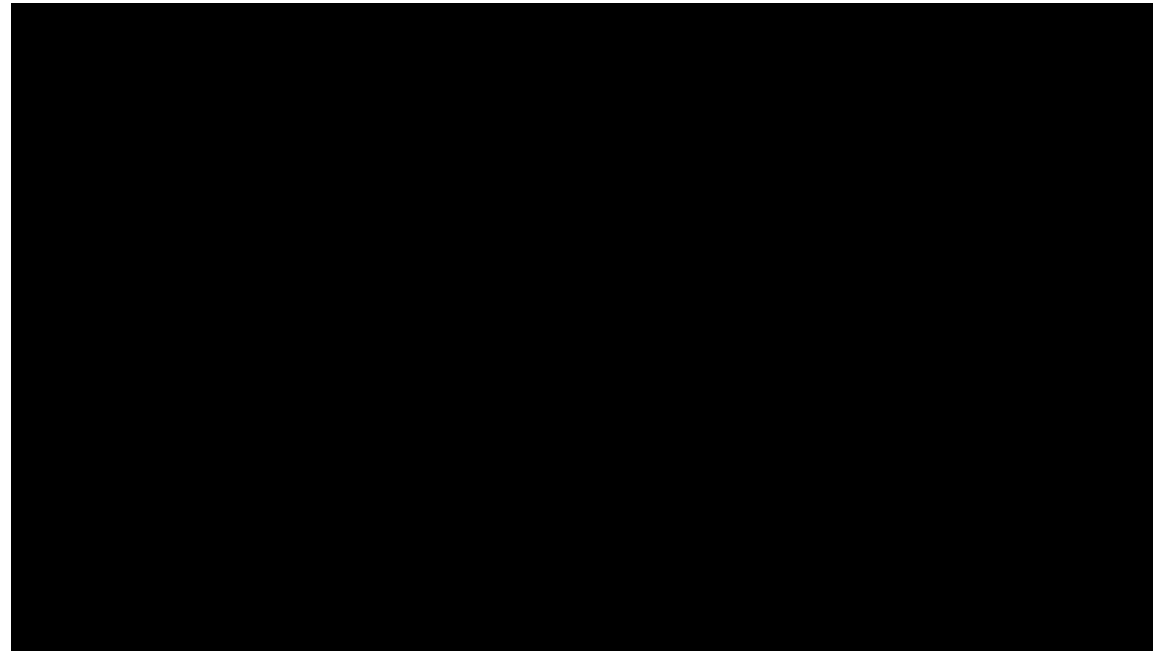
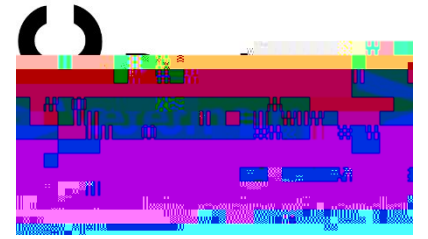




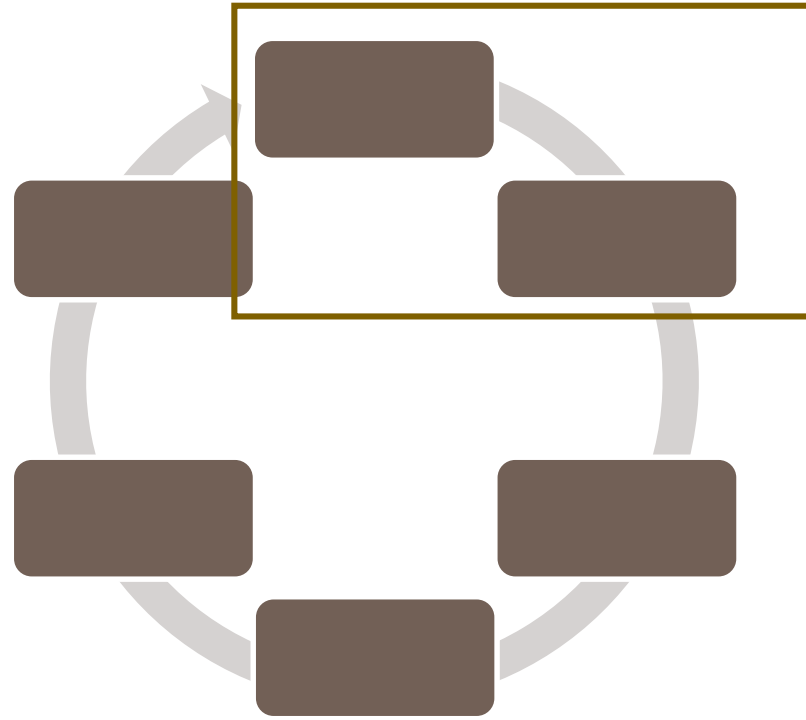
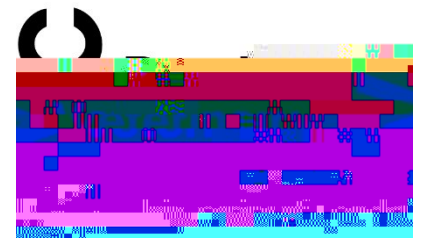
Genomic signatures to guide the use of chemotherapeutics



Excerpt from Keith Baggerly's lecture



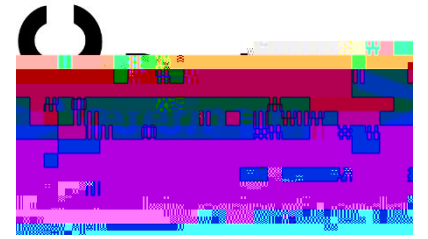
Why is RDM important?



Thinking about data collection

Methodologies and standards

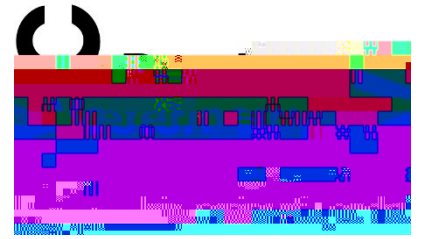
Taxonomies & controlled vocabularies



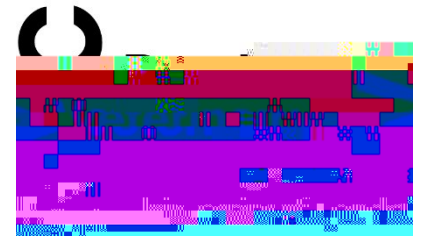
18-day pregnant females	female (phenotype)	hexaploid female	dikanvon	female, 6-8 weeks old	pseudohermaphroditic female
2-yr-old female	female (pregnant)	individual female	dioecious female	female, other	female
3-yr-old female	female	female, aged	sex female	4000-yr-old female	female child
seasonal	female	female, virgin	sexual-outspous female	adult female	female price
sterile female	gestate female	female plant	mammal female	female	female (ewmex) female
remale	female (calf)	femele	thelytoky	dioecious female	female, other

Nucleic Acids Research

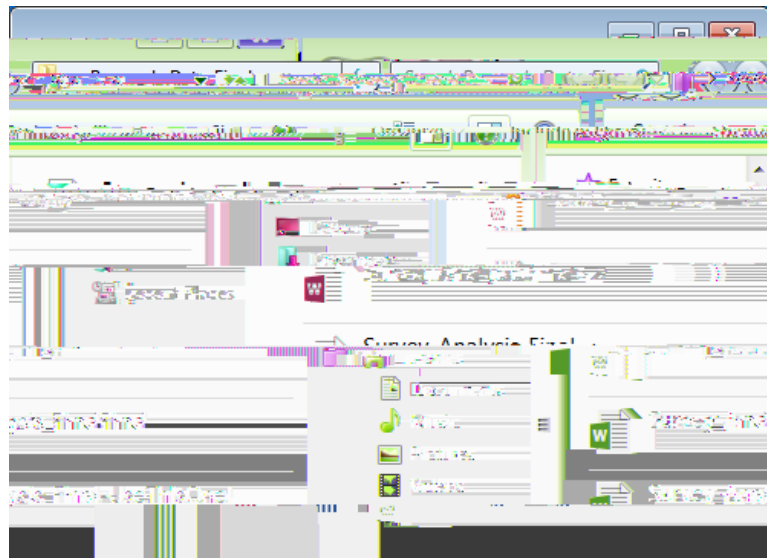
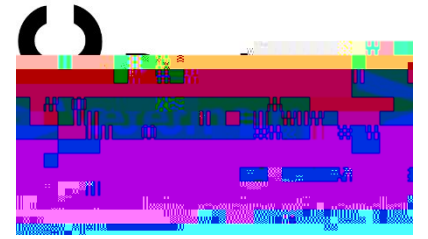
Over to you (DMP checklist)



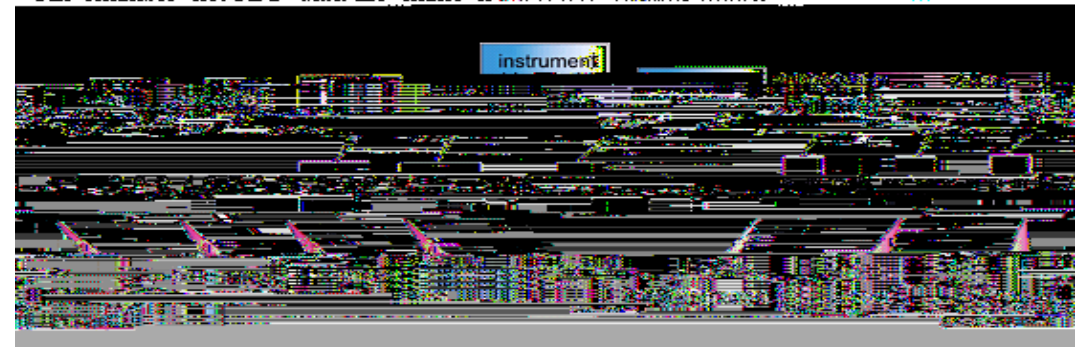
Storing, describing, depositing, citing
datasets



Naming your files

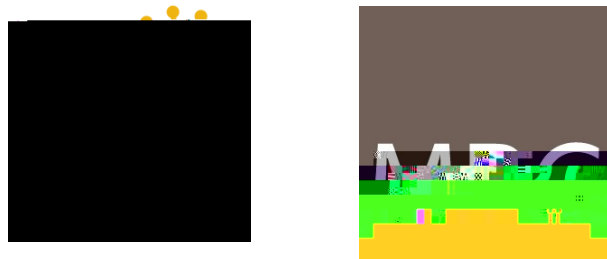
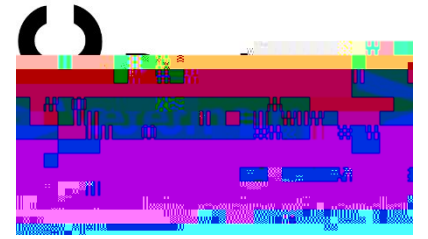


An example netCDF data file name is depicted below:



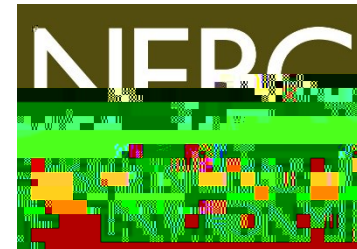
An example from LSHTM

Funder requirements



EPSRC

Engineering and Physical Sciences Research Council

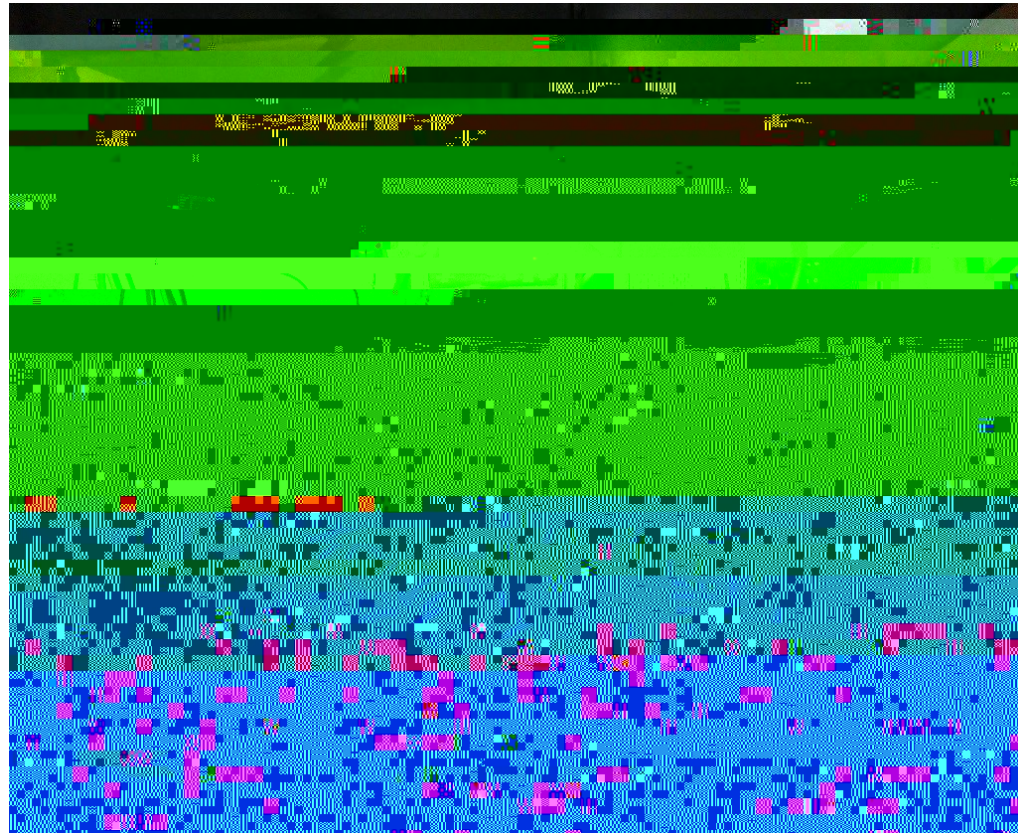
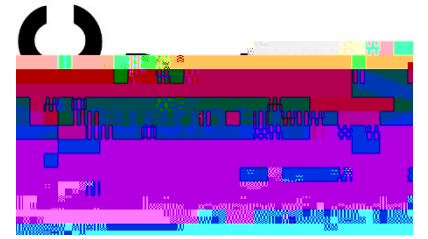


welcc

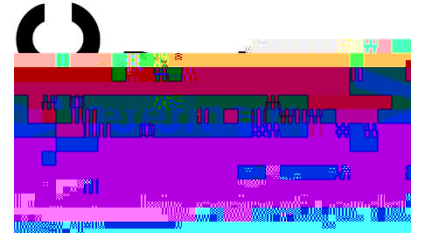
RVC RDM pol5VMCID 3BD4((abr)(3dg)(ed))]TET /P A

RV†

Good alternative to



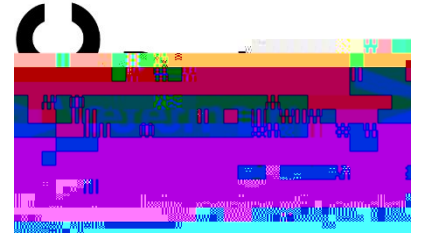
Backing up



Cloud storage



Publishing



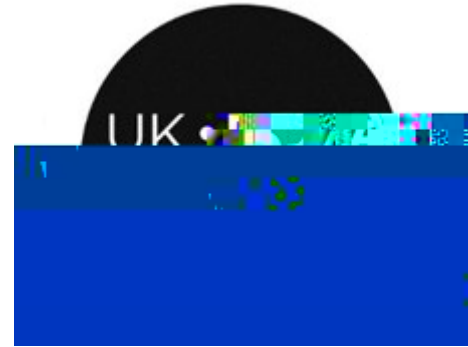
Depositing

Studies may share their data by depositing their data collection (or a subset) at a discipline based repository like the UK Data Archive (www.dasa-arve.ac.uk), or at an institutional repository that can preserve data and make them available to users.

Sequences must be submitted to EMBL Database Library or GenBank. Protein sequences that have been determined by direct sequencing of the protein must be submitted to SWISS-PROT at all accession numbers should be included in the manuscript.

We expect that all researchers submitting to EMBL (PL) (DS) (submissions) (in which) (the) software is the central

Repositories



Tools and resources



Next steps

