The metal-free capture of dinitrogen for the synthesis of azo dyes, photo-switches and flexible P,N ligands.

# General Information

## Name applicant and project number

xxxxxxxxxxxxxxxx

## Name of data management support staff consulted during the preparation of this plan and date of consultation.

xxxxxxxxxxxxxxxx

# What data will be collected or produced, and what existing data will be re-used?

## Will you re-use existing data for this research?

**If yes: explain which existing data you will re-use and under which terms of use.**

No

## If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

The data will consist of experimental procedures and reaction optimisation conditions for new compounds (or new routes to established compounds where appropriate). These will be written up in electronic lab-books (currently using Microsoft Office for this), and when published will be written up fully in a Word document and saved as a pdf.

The compounds will also be characterised, so the data include spectroscopic data (NMR, UV-Vis, IR, possibly EPR) and other experimental techniques (X-ray crystallography, mass spectrometry) and computational analysis (using DFT methods). All of the raw data files will be saved, and the results written up in Word documents (and saved as pdfs) on publication.

## 1.3. How much data storage will your project require in total?

100 – 1000 GB

This is an estimate (the raw data files for X-ray structures and DFT computations can get quite large, so the final amount will depend on how much of each of those techniques are used in the end), but the University of Birmingham Research Data Store currently provides up to 3 TB for free, which will be more than enough.

# What metadata and documentation will accompany the data?

## Indicate what documentation will accompany the data.

All of the experimental procedures will be recorded as they are being carried out, with all necessary information for the experiments to be reproduced by other researchers. These include, but are not limited to: quantities of reagents, volumes of solvent, size of glassware/reaction vessel used, wavelength of light, temperature of reaction, length of time, characterisation techniques etc, and the summaries will be saved as .ppt, .doc or .pdf files, as appropriate. The experiments will be recorded in electronic labbooks (saved as .ppt files), so they can be easily searched for keywords, reagents or other search terms.

*Metadata for data deposited in the UBIRA eData Repository will use Dublin Core as metadata standard.*

When the results are published, they will be written up entirely in the experimental section (or supplementary information, depending on the journal in question) for other researchers to use. CIFs (data files for X-ray crystallography) will be included with the accompanying manuscript on submission for publication.

## Indicate which metadata will be provided to help others identify and discover the data.

The data will be stored in readily searchable folders/directories. The folders/directories will named according to a standardised format within the group, with brief but descriptive titles that are the same for each researcher. Newer versions of documents will be saved with the version number and date appended to the title for clarity (e.g. paper1\_v1\_YYYYMMDD.doc). The electronic labbooks will clearly display the name and location of any experimental (characterisation) data related to that particular experiment, so that cross-referencing and future re-use is as simple as possible.

Depositing data at the Ubira eData repository would follow Dublin Core as a metadata standard and the minimum metadata provided for published datasets will cover amongst others title, type of data, creators, publication date and related publications. *As a DataCite DOI will be minted to data deposited in the eData Repository, the dataset will be discoverable through the DataCite data search and other aggregators. In addition, the repository will be indexed by search engines such as Google* making the data available to the broadest possible audience*.*

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

## Describe where the data and metadata will be stored and backed up during the project.

Institution networked research storage

During the project, all raw data will be stored and backed up in the in BEAR (Birmingham Environment for Academic Research) Research Data Store, for five years, and any data (raw or processed) further used for scientific studies will be archived on the UoB Bear Archive. I will also make use of the UoB Open Access repository Ubira eDAta, which is available over the internet and indexed by search engines including Google Scholar and OpenDOAR, thereby making the data available to the widest possible audience.

## How will data security and protection of sensitive data be taken care of during the research?

Default security measures of the institution networked research storage

The data will all be related to chemical compounds that are synthesised, and so no personal data will be generated/collected.

# How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

## Will you process and/or store personal data during your project?

**If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?**

No

No personal or sensitive data will be generated.

## How will ownership of the data and intellectual property rights to the data be managed?

The data produced in this project will be of significant public interest. These include the spectroscopic data, experimental procedures and reaction optimisation conditions for new compounds (or new routes to established compounds where appropriate). There is the strong potential for parts of the generated data to be of commercial interest, in particular the sustainable generation of diazonium salts. I will therefore maintain a professional management of intellectual property rights, guided by the expertise of the Birmingham Enterprise office at the University of Birmingham. Non-patented results will be disseminated rapidly through peer reviewed (open access) journals, at seminars and at conferences.

# How and when will data be shared and preserved for the long term?

## How will data be selected for long-term preservation?

All data resulting from the project will be preserved for at least 10 years No data will need to be destroyed as none of it will be personal.

## Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

**If yes, please explain.**

No

In general there are not any reasons for restricted access.

However, if the results are patented then there may be some restrictions outlined in those documents, but I will receive guidance on this process from the UoB Enterprise team if/when this is necessary. Results that are published in papers will be be publicly available, notwithstanding the above statement about IP.

## What data will be made available for re-use?

All data resulting from the project will be made available

All data resulting from the project that underpin publications will be made available for re-use, notwithstanding IP considerations as advised by the UoB Enterprise team. I will certainly be following the "As open as possible, as closed as necessary" principle.

## When will the data be available for re-use, and for how long will the data be available?

Data available as soon as article is published

The data will be made publicly available on publication in journals of the corresponding work.

## In which repository will the data be archived and made available for re-use, and under which license?

All raw data will be initially stored on the University’s Research Data Store for five years, and any data (raw or processed) further used for scientific studies will be archived on the University’s Bear Archive. I will also make use of the University’s Open Access repository, which is available over the internet and indexed by search engines including Google Scholar and OpenDOAR. In addition, a metadata record will be created in the University’s research information system PURE and the dataset is thus also discoverable through the University of Birmingham Research Portal (https://research.birmingham.ac.uk/portal/en/) under an open licence.

## Describe your strategy for publishing the analysis software that will be generated in this project.

The published data will be written up in manuscripts that can be downloaded as pdfs. Images of the spectra will be included in the supporting information, which is also saved as a pdf for ease of use. The raw data will be saved in the format that it is produced, and in certain cases, there will be specialist software required to process that data (e.g. for NMR spectra, the .fid files are processed using TopSpin, which is freely available from Bruker), but the results will be accessible to all.

# Data management costs

## What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The University of Birmingham Research Data Storage currently allocates up to 3 TB of data for free, which will be more than enough to store the data from this project. Deposited data in the Ubira eData repository will be reviewed by authorised staff from the Scholarly Communications team, making sure that all published data in the supporting information are FAIR.