

## Description Objectives Ethics Resources (DOER) Document for Molecular Structure Drawing Prediction

### Problem Description:

Researchers in Chemistry, Physics, or Biology, produce large amounts of data “big data” which is difficult to manage and search. The School of Chemistry at the University of St Andrews generates this data on a daily basis. NOMAD, an online NMR data management tool, is used to store this data.

I will be looking into the possibility of creating a drawing tool for molecular structures that makes predictions on what a user is trying to create. The tool would use the data stored in the NOMAD database to make these predictions in real time as the user draws. It would store a data structure for the molecular structure being build and compare it with past built structures to make its predictions. This project's main aim is to try and improve the efficiency of drawing a molecular structure.

The project would be built outwidth the NOMAD, with a defined API for the NOMAD project to communicate with at a later date. The data from the NOMAD database would also be extracted for testing purposes only.

### Objectives:

#### *Primary*

- A tool that makes suggestions on a molecular diagram as it is being drawn.
- The tool should be able to get its prediction correct, with one component on the diagram left to draw around 50% of the time.
- Good user interface for suggestions for testing.

#### *Secondary*

- The tool should be able to get its prediction correct, with one component on the diagram missing almost every time given it has been drawn before.
- Make good educated suggestions with more than one component missing.
- Tool should be decoupled from the interface and database, so that it can easily be used with any interface and database.

#### *Tertiary*

- The tool learns from users diagrams that it doesn't get correct or hasn't seen yet and improves in its suggestions as its database grows.

### Ethics:

This will not involve a controlled experiment.

The data being used from the NOMAD database will come from the St Andrews Chemistry department but after use will be destroyed.

I will gather information from potential users in the St Andrews Chemistry department hopefully. I will film their hands only when drawing structures or just take a recording of a user drawing the structure on the computer. The video footage that will be obtained from

lectures or students drawing structures but will only be used for this project, viewed only by myself and destroyed after the project has ended.

**Resources:**

- Video camera.
  - Used to film chemistry students and lectures drawing molecular structures.
- Subjects.
  - Gathering data on drawing molecular structures.
- School lab machines.
  - research and development.
- Access to the NOMAD project.
  - To take data from database and put into my project.
- Version Control Systems (Mercurial)
  - That is what Nomad uses.