**Digital Records Curation Programme (DRCP) - English Version**

**Week 3 – Workshop: File Profiling**

**Learning Outcomes**

At the end of this class, students should be able to:

- explain what file profiling is and why we do it
- understand the concepts of representation information
- understand checksums and technical registries
- use file profiling tools at a basic level

**Lesson Plan**

**1. Lecture – File Profiling (10 – 20 minutes)**

Use the DRCP slides for this class to provide a short lecture on file profiling.

**2. Using DROID**

Lead the students through the use of DROID, covering:

- installing DROID
- running one file through DROID (pointing out and explaining PUIDs and MIME types)
- running one folder or group of files through DROID
- filtering results
- running reports (give students time to explore different kinds of reports)
- asking students to read section 3.11 on checksums in the DROID user guide: [http://www.nationalarchives.gov.uk/documents/information-management/droid-user-guide.pdf](http://www.nationalarchives.gov.uk/documents/information-management/droid-user-guide.pdf)
- generating checksums (go to Preferences – Profile defaults – tick Generate hash for each file using SHA-1 (This defaults to unticked because generating the checksum slows DROID down))
- running one file to see an example of a checksum
- checking for duplicates by running one folder or group of files with checksums, exporting the results as a CSV file, opening the file in Excel or similar spreadsheet software, highlighting the checksum column, selecting Home Section – Conditional Formatting – Highlight Cell Rules – Duplicate Values. This will

highlight all duplicate checksums. This is an opportunity to explain how DROID can help with digital records appraisal, and also to explain the concept of meaningful duplication.

## 3. Using JHOVE

Lead the students through the use of JHOVE (pronounced Jove), covering:

- installing JHOVE
- the differences between the command line interface and the graphical user interface
- running a file through JHOVE
- running a folder or group of files through JHOVE (drag and drop)
- generating checksums

Explain that DROID and JHOVE do similar things, but there are differences between the programs, and the students may want to use them both when they are file profiling. For instance, DROID may identify a wider range of file formats, but JHOVE will provide richer metadata.

## 4. Conclusion

Summarise what you have covered in class in relation to learning outcomes for this class.

Be sure to relate file profiling to the concept of 'identity' and checksums to the concept of 'integrity' covered in the second class.