

Sensitive content

In June 2010 UCL Library Services conducted a sector-wide survey into thesis deposit and open access in UK universities and HEIs. A rich body of data was provided by respondents describing how they approach the issues surrounding open access deposit of electronic theses. The data was analysed and evidence is presented here of consensus and best practice. A survey summary report and full text responses are available on the project website¹.

Of 144 survey respondents, 75 (52%) report concerns about sensitive data within theses. Such concerns are equally divided between students, supervisors and administrators:

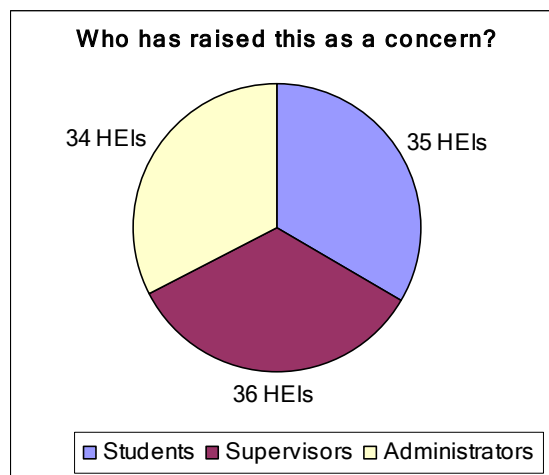


Figure 1: Sensitive content: Number of HEIs where concern raised

Only 3 respondents (2% of HEIs) reported this as a frequent or top concern:

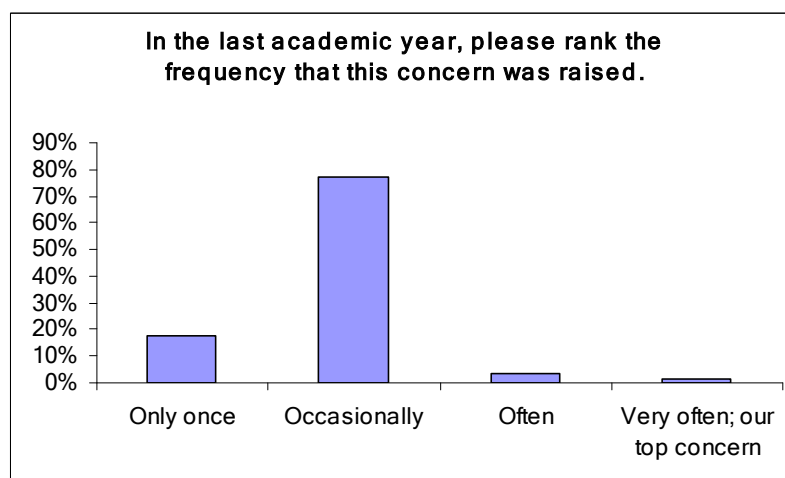


Figure 2: Sensitive Content: Frequency of concern

¹ Project website <http://www.ucl.ac.uk/ls/etheses>

Examples of concerns raised

The nature of sensitivity is varied, although commercial sensitivity was most commonly cited:

- “Commercially sensitive information”
- “Patentable knowledge has been made accessible preventing it from being patented in the past”
- “Agreements negotiated by Research and Enterprise Dept for use of commercial company data normally require a minimum of ten years guaranteed confidentiality”

Political sensitivity, covering a wider range of issues, was also a significant cause for concern:

- “Author considered life would be at risk”
- “Confidential government statistics”
- “Intelligence and security theses”
- “One student raised the concern of their work (an anthropological study of guest workers in [middle eastern country]) being used to identify and target vulnerable communities and people, because of its wide dissemination”

Other concerns cited tended to derive from the nature of the research process or its subject matter:

- “Theses have contained data that is a) relevant to court case, b) contains animal experimentation, or c) contains information on child welfare cases that could be identified in the field, even though anonymised.”
- “Students inform us that participants in (eg. case study) theses often only agree to be interviewed if confidentiality is guaranteed”
- “Health theses contain patient information”
- “History thesis which included names of supporters of Fascist organisations”

Mitigating actions

Respondents were invited to describe in their own words what actions they take to address these concerns. Two are still formulating their procedures and one said the issue has not yet arisen. Of the measures taken to address concerns, notable examples are discussed below.

The vast majority of respondents (79%) restrict access to the thesis, 47% via embargo and 32% via other restrictions e.g. opt-out altogether. Only 17% advocate the creation of a different or redacted version for public access. Instead, 25% use education and training or expect candidates to omit sensitive data from their theses.

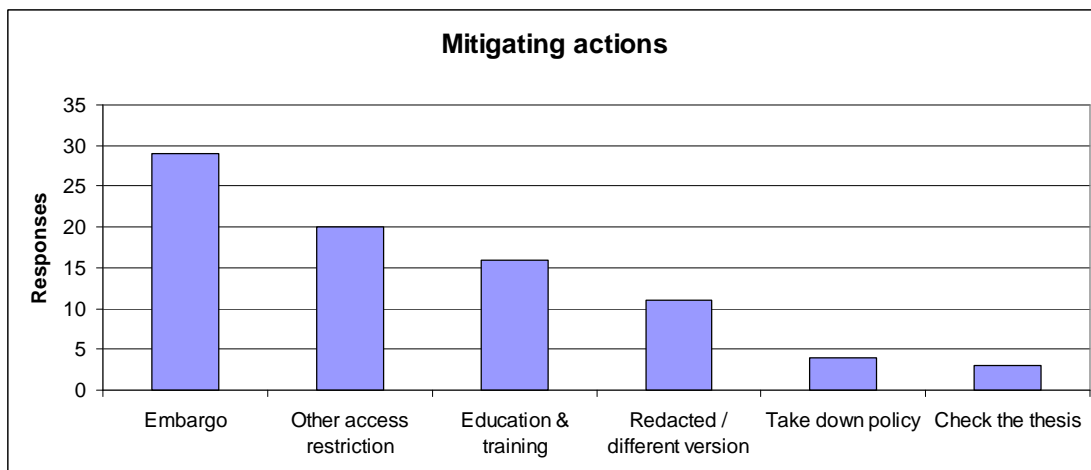


Figure 3: Sensitive content: Mitigating actions

Embargo

The embargo system is the most popular method of addressing issues of data sensitivity, in operation at 29 HEIs.

- “Students have the right to embargo their thesis for this reason. Risk management process in operation”
- “Robust embargo procedure. Partial embargo to remove really sensitive information.”
- “Respect embargo requests”
- “Candidates are given the opportunity to embargo their theses in consultation with their supervisors”
- “Confidentiality issues are covered by the embargo procedures. Before any embargo is taken off and the e-theses uploaded we look at the reasons for embargo and engage with the Graduate School to see if they want an extension on the time period”

- “In one case, the student was granted an embargo by the university research committee (after following the university's normal process for requesting confidentiality).”

Other access restriction

In addition to embargo restrictions above, a further 11 HEIs limit access to the thesis and a further 9 would not put the thesis into a repository at all.

- “Where such materials exist and permission to make the material open access can not be got, then the final version remains closed access”
- “Again, this is covered by the restricting access to your thesis form.”
- “If any of the above applied then we would not upload the thesis.”
- “Restrict access to online theses”
- “Student opt out”
- “These can be Restricted or Confidential for specified amount of time (specified by author and supervisor). If anyone wished to consult these they would need special permission by letter from the Head of the Department that the thesis originated from.
- “Vols containing such info retained by research office. Would not be deposited in repository.”
- “We can restrict any e version in same way as print if this is a concern.”

Education and training:

Sixteen responses revolve around education and teaching good academic practice.

- “Emphasise that students should be following good practice in these areas in any case (regardless of electronic access).”
- “The Graduate School gives guidance on inclusion of sensitive data etc - anonymising, restricted access”
- “Guidance notes.”
- “Students receive instructions regarding omission of confidential/sensitive material”
- “Removing sensitive material to an appendix and embargoing that separately so that the main body of research is still available publicly”

- “Students should notify their subjects that the output from their study may be placed in a digital archive, which is available globally.”

This respondent implies that embargo solutions are only available to theses written or started under previous regulations and that the expectation now is for correct treatment of data to be built into research practice:

- “Training for new students to make them aware of what permissions they need to request. Training for existing students and supervisors about the use of embargoes”

Redacted / different version:

Eleven respondents allow for redaction, i.e. the blanking out of sensitive data, or for separate public access versions to be created. This may be seen as a last resort measure or used in conjunction with other measures:

- “Takedown policy, redaction, embargo. These mitigations may be subject to approval by Research Degrees Committee “
- “Embargo option is available or blanking out sensitive information. A special text reference can replace sections of the PhD that are confidential etc.”
- [where permission is sought but not obtained] “the student has the opportunity to produce a redacted version.”
- [in a specific instance] “We agreed to remove any references to particular people and to not include several appendices.”
- “There may be a case for producing a version of the thesis for open access, i.e. one that has such material excluded but that can be provided on request.

Take down policy:

Four respondents use their take-down policy in cases of mis-exposure of sensitive data. This is a passive response to managing the issue, however, and institutions should take active measures to avoid this becoming necessary. A take down policy works best as the final part of a combined approach to managing this risk:

- “Robust research training and a clear takedown policy.”

Check the thesis:

Only three HEIs mention the resource-intensive task of checking the thesis.

- “Explain that theses are checked to make sure that there is no sensitive material included before it is made live.”
- “Check to ensure information is anonymous”

Other measures:

One respondent notes the approach taken for ethically sensitive research:

- “Standard ethical declaration form and referral to university ethics committee when required.”

Finally, this respondent suggests that open access should be regarded as the norm for the greater part of research output:

- “Research inappropriate if none of it can be made publicly available or not available at all”

Actual occurrences

When asked how many theses require embargo or redaction, only 29% of HEIs said “none”, indicating a higher incidence rate than with other concerns. Another 29% had fewer than 5 affected theses, and a further 10% up to 40. Only 2 HEIs report a higher number (300 and “a significant number”, both high volume Russell Group universities). Of those who said “don’t know”, three noted that the process is pre-library and they would therefore not receive or know about those theses.

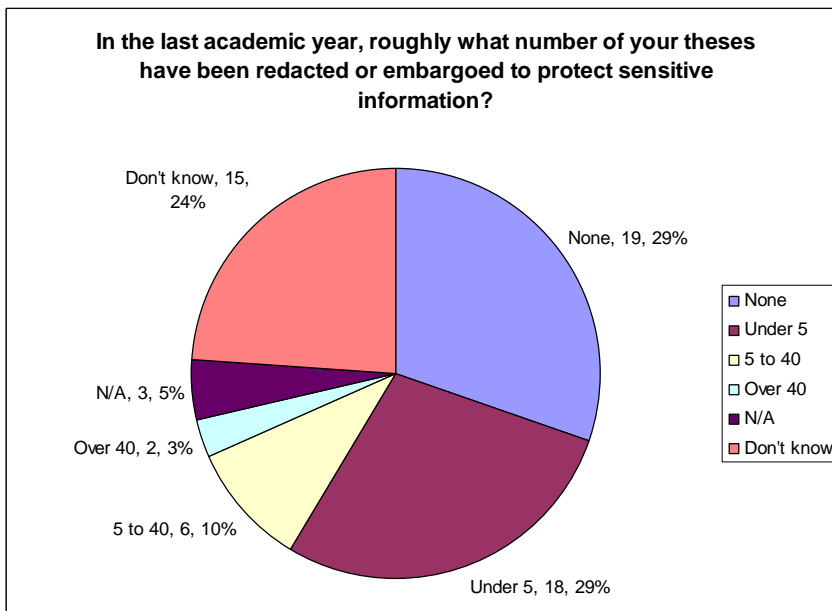


Figure 4: Sensitive content: Actual occurrences

Conclusion

The nature of doctoral research and its funding and sponsorship opportunities mean that inevitably some students' work will gather or make use of sensitive data, and most universities have measures in place for access restriction to address that eventuality. However with the addition of a programme of education to promote good practice, by omitting or anonymising confidential data, it should be possible for the majority of theses to remain intact. If sensitive data has to be included, keeping it in an appendix allows the main body of research still to be accessible.

Using this combination of measures, it can be seen that very few HEIs, even the larger, more research intensive universities, have needed to redact or embargo significant numbers of theses. This suggests that research ethics and scholarly practice training are sufficient in most cases to ensure that issues of sensitivity are not a barrier to open access.

Authors:

Josh Brown (SHERPA-LEAP) and Kathy Sadler (UCL Library Services), Project Officers on 'Influencing the Deposit of Electronic Theses in UK HE' project.

Further information can be found on the project website at <http://www.ucl.ac.uk/ls/etheses>