IPR and Licensing issues in Derived Data

Naomi Korn, Professor Charles Oppenheim and Charles Duncan

May 2007

Context

"Data that already exists that is used to do something new. A good example is text mining where thousands of papers are put into the machine and at the end you can get a précis of the most relevant ones and, potentially, links that have not previously been discovered by researchers. The big issue is that when a piece of research is written then the sources are attributed; if you have thousands of papers do you attribute them all, do you attribute according to importance in the results, etc, etc. There are numerous issues and it's currently a very grey area, meaning that if you're working in it you're never quite sure whether you're within the law."
James Farnhill, JISC Programme Manager.

Vast quantities of cross disciplinary base data are being continually generated by automated tools and human intervention, the value of which is fundamental to research within the Higher and Further education communities. Indeed, there are a myriad of projects which are discovering, delivering and using all types of data, (for example, see http://www.e-science.clrc.ac.uk/ projects/). In order to facilitate the reuse of this data and elicit meaningful patterns, tools have been developed which provide instantaneous correlation and summary of base data –including the outputs from text mining, data mining, geospatial data, the analysis of eThesis or other operations. Whilst this derived data can be reused in its core form, depending upon the terms under which it is accessed, it may then be supplemented and overlain by additional data. *For example, derived data may be taken from oceanographic studies, and enriched with researchers' own data and made available on a repository.* These activities raise complex IPR and licensing issues, including the potential creation of new IPR from multiple texts/datasets, etc., and issues associated with reuse of the data.

In response to the issues that derived data raises, a number of international initiatives and potential solutions have been developed. These include an abstract specification for the management of digital rights in the area of geospatial data and services - the Geospatial Digital Rights Management Reference Model, (GeoDRM)[1] which is offered by the Open GeoSpatial Consortium, Inc and includes representatives from Ordnance Survey.

The aim of the GeoDRM initiative is to:

- Create a conceptual model for digital rights management of geospatial resources, providing a framework and reference for more detailed specification in this area.
- A metadata model for the expression of rights that associates users to the acts that they can perform against a particular geospatial resource, and associated information used in the enforcement and granting of those rights, such as owner metadata, available rights and issuer of those rights.
- Assess the requirements that are placed on rights management systems for the enforcement of those rights.
- Examine how this is to work conceptually in the larger DRM context to assure the ubiquity of geospatial resources in the general services market.

[1] http://portal.opengeospatial.org/files/?artifact_id=17802

Scope of this study

The JISC IPR Consultancy was asked by the JISC Information Environment to carry out a short scoping report as JISC recognizes the importance and benefits of being able to reuse a combination of data to both the information environments and e-research.

This scoping report provides an analysis of the IPR and licensing issues inherent in derived data as a means to build a framework of existing practices and possible solutions. It does this by examining two case studies, i.e., the JISC-funded National Centre for Text Mining (NaCTEM)[2] and Geospatial data as well as referencing the GRADE Project[3] This exercise is also important in providing the context for the applicability of broader national and international initiatives, as above, as a means for potential future studies into their implications and applicability within HEIs and FEIs.

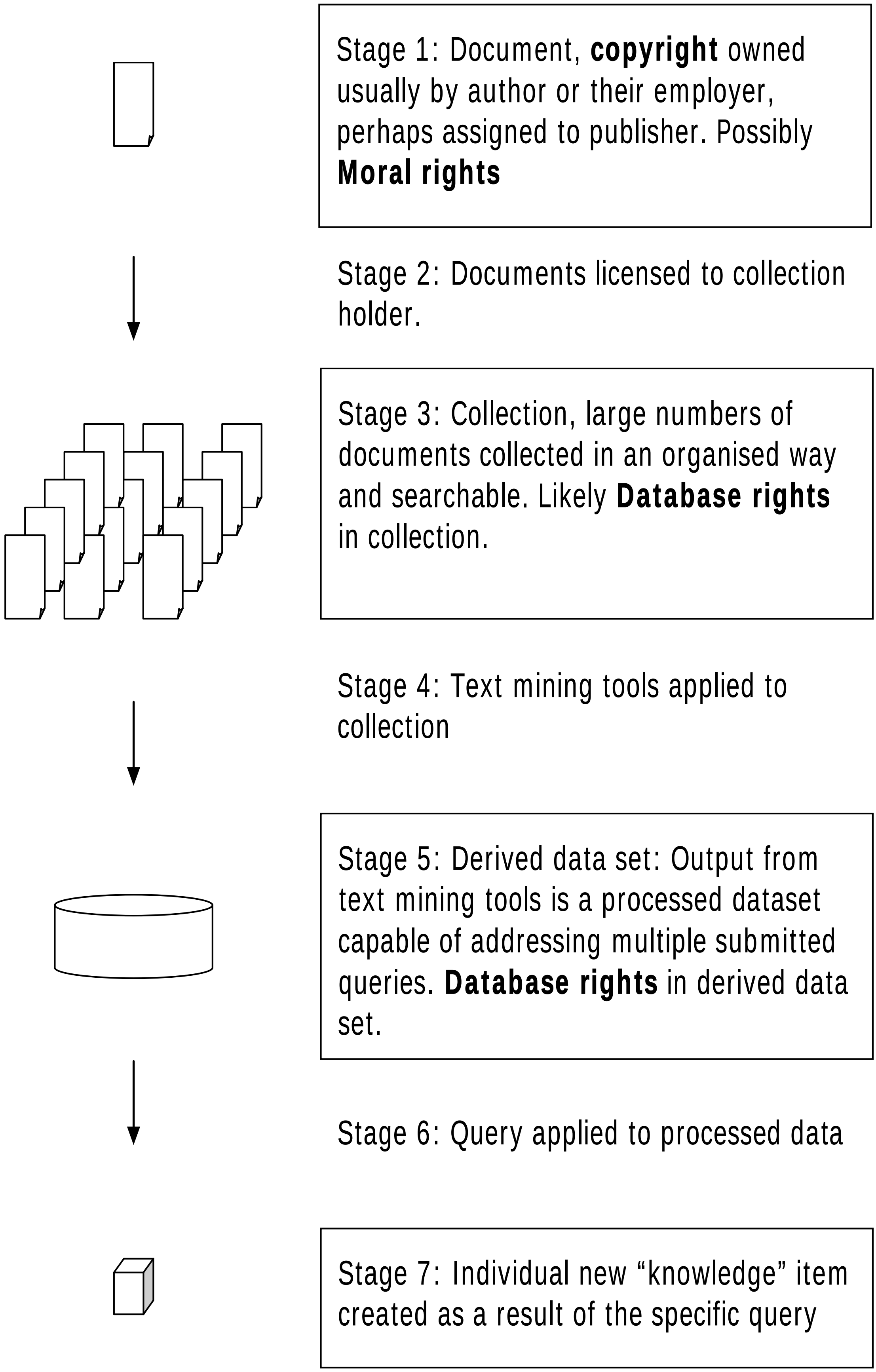
This report scopes the key IPR and licensing issues associated with text and data mining from a **UK** perspective. Legal issues associated with international jurisdictions and multiple territories, are beyond the scope of this current study.

Case Study 1: National Centre for Text Mining

"Text mining attempts to discover new, previously unknown information by applying techniques from information retrieval, natural language processing and data mining."[4] Text mining is based on the processing of large numbers of documents and results produced from text mining are a form of derived data. Text mining tools are applied to large collections of text documents such as Medline[5] (commercial service, 11 million citations) or Public Library of Science[6] (Open Access Service using Creative Commons Attribution licences), as well as applied to private collections of documents. A likely scenario may include the retrieval of chemical formulae in publications linked to chemical formula in databases and then made available for non-commercial use. Text mining tools can also be applied to any collection of documents.

Figure 1 (below) graphically represents the process of creating a derived text based dataset and the mix of IPR and licensing issues which are likely to arise as a result of the challenges arising from machine generated datasets.

[2] http://www.nactem.ac.uk
[3] http://edina.ac.uk/projects/grade
[4] http://www.nactem.ac.uk/faq.php?faq=1
[5] http://medline.cos.com/
[6] http://www.plos.org/



Stage 1: Document, **copyright** owned usually by author or their employer, perhaps assigned to publisher. Possibly **Moral rights**

Stage 2: Documents licensed to collection holder.

Stage 3: Collection, large numbers of documents collected in an organised way and searchable. Likely **Database rights** in collection.

Stage 4: Text mining tools applied to collection

Stage 5: Derived data set: Output from text mining tools is a processed dataset capable of addressing multiple submitted queries. **Database rights** in derived data set.

Stage 6: Query applied to processed data

Stage 7: Individual new “knowledge” item created as a result of the specific query

Figure 1: Stages in text mining with IPR considerations

Stage 1: At the item level, in terms of IPR, there are likely to be three key rights which exist in the original documents:

- Copyright in the original text and associated metadata which may still belong to the authors, or their employers (if created by employees under contract or, during or, on behalf of the employer), or copyright may be assigned to publishers.

- Database rights may have been afforded to the organisations who have invested efforts in collating collections or publications and making them searchable, regardless of the different types of data sources that are searched or the business models under which they are made available (such as PubMed or Medline). A database is defined in law as any collection of data, text, etc. Many databases are protected by so-called database right in EU member states. This right is afforded automatically if there has been substantial investment in verifying, obtaining (i.e., gathering) or presenting the contents of a database, as long as one or more of its makers is resident in the EEA or a body operating in an EEA state. (EEA = European Economic Area, i.e., the European Union plus a couple of other West European countries).This right prohibits extraction or re-utilization of a substantial part of the database (in terms of both quantity and quality), unless a licence has been sought or unless it is for illustration within teaching, or research purposes which are non-commercial. In all such cases, the source must be indicated. Database rights may be present within sites and services which collate and display derived text or other data, and it would be infringement of this right (and perhaps also copyright – see above) if this material were reused within another service or system without licence from the rights owner.

- Moral rights. Moral rights protect the reputation of the creator of the work. Although the moral right of attribution (to be identified as a creator) requires assertion and is not in any case available for employee-created works, or for works that appear in journals or encyclopaedias, the right to object to derogatory treatment (which will last as long as copyright and does not require assertion) may apply unless waived.

Stage 2: Permission will need to be negotiated from the collection holders to collate and analyse the original documents.

Stage 3: Database rights may arise in the course of collating and making searchable the documents, regardless of the types of material or databases that are searched.

Stage 4: At the next stage, the National Centre for Text Mining applies its tools to text collections to analyse the documents using complex natural language processing techniques. After this point, the extracted data is analysed by the data mining component which generates further data.

Stage 5: After the substantial processing work a new “derived” data set has been produced. A useful analogy in terms of IPR situation can be drawn between the results from the text mining with those arising within the index produced by web search engines. A significant licensing issue has been identified relating to the resulting output: The output is not in the form of extracts of the text identifiable from individual works but instead, the output comprises of associations and patterns whose value is related to the number and type of occurrences in large numbers of works. This process makes it impossible to cite every work used (as is usually required by the collection provider) or evaluate the contributions made by each work individually. On the other hand it may be difficult, except in isolated cases, to prove that data has been extracted from a specific work. There is also a specific type of work allowed for under copyright law, i.e., computer-generated work. There is an argument that works created following text mining might fall under that heading. Ownership of such works resides with “the person responsible for the arrangements necessary for the creation of the work", an ambiguous phrase whose meaning is unclear.

Other issues at this stage include:

- Licences, such as those from Medline, which prohibit the display of the body of the work. This is important because text mining is a sequence of processes, and usable output can be obtained from each of the processes. So, for example, some text mining tools extract information (personal names, geographical locations, company names, names of genes or proteins, etc.) and then find facts or instances of events involving these entities in a text. These results are usually shown by displaying the processed article and highlighting the discovered entities and facts.

  For example: "It may be that a user wants to curate such proposed extracted facts for accuracy, before committing them to a subsequent data mining phase. However, if we display a representation of an article with highlighted facts, it may infringe licensing restrictions. Although they are generating the display from an underlying representation into which the system has analysed the original article, it will look to the user as if we are displaying the original. In other words, there is apparently little difference between running a search query, receiving a set of articles to process, processing them and showing the highlighted facts, on the one hand, and running a search query, receiving a set of articles and displaying them. The only difference is in the highlighted facts, the text is still being retrieved and displayed, to all intents and purposes". John McNaught, Associate Director, National Centre for Text Mining

- There may not necessarily be rights from the standard licences offered by licensors to process the text by machines. For example, the Creative Commons Attribution licence used by Public Library of Science is based on human use rather than machine processing.

- Attribution to the copyright owner of a work or collection is not just normal practice in research, but is also required in the law if the copying was for non-commercial research or private study. However, if the output from a text mining exercise comprises small components from many different sources with different owners, attribution of individual pieces of information is not really possible because of the lack of identifiable individual works. In this case, perhaps the "use" of the work is much less than would require attribution, i.e., what has been copied is a non-substantial part of the original work, is therefore not copyright infringement, and therefore no permission or attribution is required.

- Who owns the rights in the resulting dataset? The employer of the person who carried out the text mining, or the individual him/herself may well own the rights alone, or jointly with the owners of rights in the original material from which the dataset derived. The situation is particularly complex and depends on the circumstances; judgements can only be made on a case by case basis.

- In any case, what are the relevant rights? Is this newly created data subject to copyright and/or database rights? Again, judgements can only be made on a case by case basis. Data derived from text mining could be considered a form of sophisticated metadata.[7]

Stage 6: The next stage of text mining is to submit queries to the derived data set. These queries may be generated by the owner of the dataset or they may be offered as a service by the owner so that users can query the dataset and recover the results with no one else being involved.

Stage 7: The resulting "knowledge" items may have considerable value as the patterns or associations uncovered may not have been deduced by any other means.

[7] See for example http://www.jisclegal.ac.uk/publications/ethesesandrew.htm which suggests that metadata and electronic theses may demand protection as a skilled piece of work.

This may be similar to a user querying a web search engine. Only that user sees and uses the results of the query. In this case the IPR is perhaps clearer as the user who generates queries is likely either to be the owner of the derived data set or will have agreed to use the derived data set under specific licence conditions.

Other additional IPR and licensing issues that can arise in text mining also include:

- Trademarks: Ensuring that any new service does not infringe the trademark of any existing service. As Trademarks are monopoly rights, trademarks can unknowingly be infringed without due care.
- Where there are several collaborators (e.g. multiple document collections are processed each with different owners) how is ownership of the derived data determined?
- There is often, within a research project, a period in which the derived data is kept private to be used only by the research collaborators. Later they may agree to the data being made more openly available for others to use. Is there a suitable "open" licence to use at this stage (e.g. Science Commons)?

Arguably, the key IPR uncertainty in text mining surrounds the inability to attribute every copyright owner/author, due partly to the vast number of articles searched but also because the extent of the copying of each article is difficult to audit, and in most – but not all - cases is probably "insubstantial" and therefore does not raise IPR issues. It is worth noting that BioMed Central specifically encourages text mining from its database[8]. However the licence terms require attribution of the authors when derivative works are created. *Nature* also encourages text mining but has proved an alternative[9] to full text access which provides "snippets" but impedes access to the full structured content – one of the key assets of text mining. The issue is well summed up by Lynch[10], who states:

> "We need license terms that minimize or render moot the uncertainties surrounding the creation of derivative works and possibly even the requirements of attribution for source materials that have contributed to the production of these derivative works."

Case Study 2: Geospatial data and Licensing

This case study relies heavily on outputs of the GRADE project and in particular the reports "Use Case Compendium of Derived Geospatial Data[11]" and "Designing a licensing strategy for sharing and reuse of geospatial data in the academic sector[12]".

Geospatial data is data related to the location of features on Earth. The features may involve topographic, geophysical, cultural, imagery or many other forms of data. The involvement of many different data sets is now a common characteristic of research and teaching involving geospatial data.

Figure 2 illustrates typical activity involving geospatial data, in which a number of IPR and licensing issues will be present.

[8] http://www.biomedcentral.com/info/about/datamining/
[9] http://blogs.nature.com/wp/nascent/2006/04/open_text_mining_interface_1.html
[10] http://www.cni.org/staff/cliffpubs/OpenComputation.htm
[11] Author Mike J Smith, http://edina.ac.uk/projects/grade/usecasecompendium.pdf
[12] Author Charlotte Waelde and Mags McGinley, http://edina.ac.uk/projects/grade/gradeDigitalRightsIssues.pdf

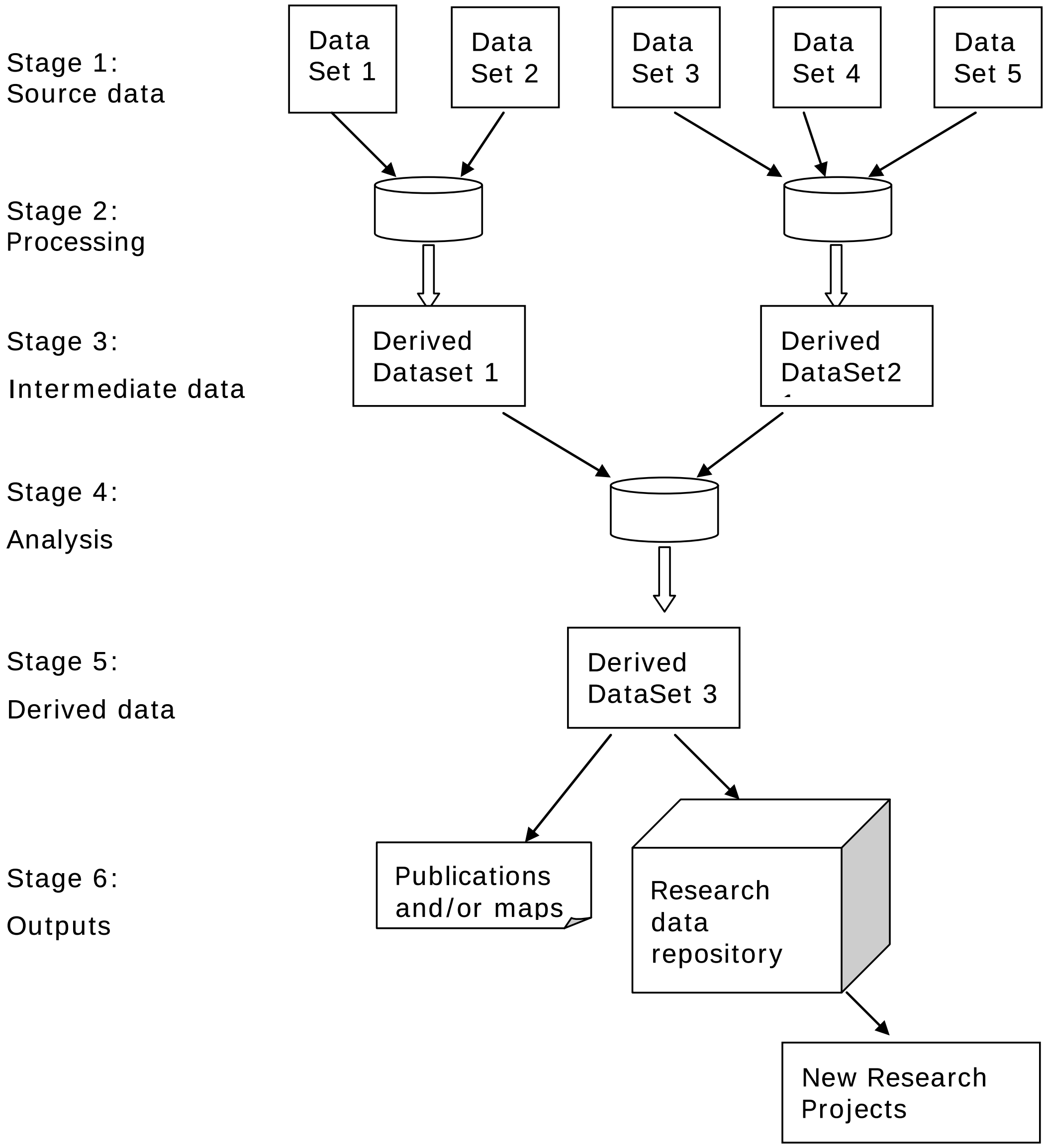


Figure 2: Typical activity involving geospatial data

Stage 1: At the top level is the identification of appropriate data sets. In terms of IPR, there are likely to be some key IPR issues which will require identification in order to facilitate the negotiation of licence agreements for the use of the data. These IPR issues, as in text mining will include the following:

- Copyright in the original data which is likely to belong to a variety of sources, although unlike text, it may well be uncertain as to whether the data itself includes enough originality for it to be afforded copyright protection. For example, in some cases, such as standard computer code and facts, the material itself may be so generic as not to be eligible for copyright protection. Case law is sparse in this area and differences between jurisdictions will also add to the complexities surrounding this issue. Even if there is doubt about whether the data is protected by copyright, almost certainly, Database Rights may have been afforded to the organisations who have invested efforts in collating collections or publications and making them searchable. Notably, the recent GRADE report[13] considered Database Rights in relation to geospatial data and concludes that: "A collection of geospatial data falls under the definition of a database in the Database Directive. No database copyright subsists in the structure of a geospatial database. A geospatial database exhibits the

[13] Author Charlotte Waelde and Mags McGinley, http://edina.ac.uk/projects/grade/gradeDigitalRightsIssues.pdf

requisite investment in obtaining, verification and presentation of the contents for the sui generis right to subsist."

The GRADE report continues to conclude that: Lawful users of the database may extract and freely reuse insubstantial parts, and anything that amounts to a substantial part may be used for the purposes of non-commercial research or illustration for teaching. However, it also observes that in the final paragraph that the greatest uncertainty is in what amounts to a substantial part!

In terms of licences, these may vary considerably and include the following:

- Ordnance Survey (OS) mapping data being made available either through the JISC OS licence[14] or specific negotiation.
- Open Access licensed material provided by researcher/institution/organisation owning the data, for example field measurements
- Open GeoSpatial Consortium Licence Agreements[15]
- Commercially sourced images with specific licence negotiations
- US satellite-based images, such as those made available via NASA[16] which are usually copyright-free and/or after the payment of a nominal fee.
- Other Crown copyright material available under a "Click Use" licences[17] or Value Added Licences[18]
- Some data, such as digital elevation datasets, may be regarded as facts and therefore not satisfying the criteria for copyright protection.

It is likely that conditions imposed by these licences will be influenced by whether the end purpose is non-commercial or commercial (**Stage 6**). It is clear that there are likely to be many different types of licence conditions imposed on the source data, and so it is imperative that suitable systems are employed to monitor these uses.

Stage 2: Processing geospatial data, like text mining, usually results in at least one derived data set. This processing stage may require advanced approaches, for example to merge datasets of different resolutions or data types (vector and raster), or to analyse data using techniques, both quantitative and qualitative.

Stage 3: Like text mining, the nature of the derived data makes it extremely difficult to determine the individual contributions of source data to the derived data and hence will make identification of the original copyright holder near impossible.

Stages 4 and 5: The greater the number of stages which process or analysis the data, the increased likelihood as to the complexity in identifying the original data components in the resulting derived datasets. While the processing activity is almost always a skilled piece of work which could justify the definition of a new work, the licence conditions of the source data sets will be inherited. Since the conditions of each licence will vary, the conditions of the most restrictive licence will take precedence. However, there may also be situations in which licence conditions conflict, for example whilst OS licences require no redistribution, other data sets use "viral" licences such as GPL or Creative Commons share-alike licences which require derived data to be shared under the same conditions as the source data.

A further difficulty arises because the duration of source data licences may be different. Some licences may be perpetual while others may be time limited. This

[14] http://www.jisc-collections.ac.uk/catalogue/coll_digimap/coll_digimapsub.aspx

[15] http://portal.opengeospatial.org/modules/admin/license_agreement.php?suppressHeaders=0&access_license_id=3&target=http://portal.opengeospatial.org/files/index.php?artifact_id=17802

[16] http://www.nasa.gov/multimedia/guidelines/index.html

[17] http://www.opsi.gov.uk/click-use/index.htm

[18] http://www.opsi.gov.uk/click-use/value-added-licence-information/index.htm

suggests that derived data would have to be withdrawn when the licence with the shortest duration expires.

Stage 6: Outputs from derived data sets can take many different forms. It is perhaps worth drawing a distinction between data and maps. Maps are one of the outputs that can be produced from the derived data. Maps may also be part of other research publications. Increasingly, research funding bodies are requiring that data sets produced using their funds should be placed in data repositories to be used by other researchers. As a result, another significant output of a derived data set could be new research using the derived data as source data.

Many of these outputs are inhibited or prevented by current licence conditions inherited from perhaps only one of the source data sets. In particular, the licence[19] under which Ordnance Survey data is made available to the HE and FE community prohibits any redistribution and limits the ground area and map area of some outputs making it difficult to include OS data in national or regional data sets. It is worth noting that where new services have been built on Ordnance Survey data, as in the case of Digimap, Go-Geo or GeoCrosswalk specific add-on licence agreements have been established to permit these uses. Some IP owners have understandable concerns over redistribution of their geospatial data. At present, the Ordnance Survey Licence restricts the dissemination of outputs to each licensing institution, while the JISC model licence permits redistribution through repositories such as Jorum.

Concluding remarks

This study has identified some of the key IPR issues associated with derived data, and clearly demonstrates that whilst there are issues and concepts relating to IPR and licensing that can applied to both text and data, there are areas of variability which need to be identified and embraced relating to:

- The unique nature of data and text
- The priorities and concerns of the organisations for whom permissions need to be sought
- The levels of adaptation and processing
- The intended use of the outcomes

There are also a number of points arising from this study which are emphasised below.

Because copyright and Database Rights will arise within the base data (and sometimes thereafter within the derived data), there is a complex array of rights issues and licensing concerns which need to be unpicked and appropriate strategies put in place as the base data is processed and moves downstream. This, of course, includes the requirement for licensing agreements which need to be secured with rights holders for further reuse of the derived data, allowing things such as enhancement with additional data, depositing in repositories and sharing, as well as agreements with users that permit them to do these things. This need to ensure that permissions are secured with rights holders in order to ensure that data and text can be collated, processed, correlated, added to, shared and deposited is crucial.

Apart from ensuring that rights are secured by appropriate agreements, the range of licences in place and their compatibility with each other and with the resulting rights embodied in the derived data can also cause problems. Whilst in certain cases, a risk managed approach may preclude the need for certain licences, particularly in cases where it is arguable that the resulting derived data may not actually involve any rights for which permission is needed, some suppliers of data, such as the Ordnance

[19] http://www.jisc-collections.ac.uk/catalogue/coll_digimap/coll_digimapsub.aspx

Survey are likely to insist upon onerous terms. These terms will restrict the ways data can be subsequently used, irrespective of whether other licences for other data in the collection signed with other suppliers offer terms that may be far less restrictive.

In order to help provide some solutions to the complexities of the IPR and licensing issues as outlined, this report makes a series of recommendations pitched at key stakeholders.

Recommendations

General recommendations for the community

1. Academics and researchers should be encouraged to use the *JISC Licence to Publish* when publishing papers are based on primary or secondary data. This will enable the sharing of their primary data by preventing their rights from being assigned to Publishers, who may then insist upon restrictive licence terms for the use and processing of their primary data.

2. Academics and researchers should be encouraged to deposit both raw data and papers containing such data in repositories as - XML, or at least something from which XML can be derived, rather than PDFs, to ensure that text and data can be mined.

3. HEIs, FEIs and research funders should be encouraged to develop IPR Policies, with assistance and support from JISC (see Recommendation for JISC below) which refer to derived data and include the need to renegotiate appropriate rights to derived data with the original supplier, if used for publication

Recommendations for JISC

4. JISC, perhaps with international partners, should commission an in-depth study of IPR and licensing of derived data, examining a number of case studies and different licensing models across international borders. This study might examine in depth the potential role of Creative Commons and Science Commons within the use and creation of derived data. The study could then provide a framework and code of practice. This work should involve JISC Legal, and perhaps could be carried out in collaboration with SURF and the OAKLAW project to build upon the GRADE work and should result in a short briefing guide for the community on IPR, derived data and how to manage these rights.

5. JISC should play a role in developing suitable guidelines, templates and model clauses relating to IPR in derived data, for use and adaptation by the HEI and FEI community.

6. JISC should continue to negotiate with publishers for open access models and encourage them to work towards other business models relevant to derived data and text mining.

7. JISC should explore with the RELI (Registry for Electronic Licences) how far the coverage of licences for text mining and derived data might be feed into the project.

8. JISC should examine the implications and issues arising from the Geospatial Digital Rights Management Reference Model and its applicability within HEIs and FEIs, and in particular its compatibility with the six stage model outlined in the Intrallect paper on DRM[20].

9. JISC should consider the development of resources for JISC-funded projects and possibly the wider HE/FE community, in partnership with research funders and

[20] http://www.intrallect.com/drm-study/DRMFinalReportv2.pdf

others, to support the understanding and awareness of IPR issues in derived data. These might be developed in conjunction with JISC Legal and include guidelines (perhaps in the form of wikis) and workshops.

10. Netskills should be asked to develop training programmes to inform the community about IPR issues that arise in connection with derived data.

11. JISC should develop and/or make available tools to make the conversion of papers and/or data into XML straightforward.

12. JISC should explore the possibility of developing a validation tool that automatically compares deposit and use licences and ensures they are consistent.

The following recommendations relate to the creation of an IPR and Licensing Framework for JISC- funded projects:

13. JISC-funded projects should ensure that their IPR statements within their project proposals should reflect the policy/ies of the host organisations involved in the project, as well as being in line with JISC's conditions of funding. IPR statements should also be used to state under what terms (or specific type of licence) users might access processed data, as well as addressing whether the deliverables will be accessible under Open Source/ Open Access principles.

14. Derived data arising from projects are likely to have a number of IPR issues associated with them. In particular, this will include copyright and possibly database rights. The code of any systems will be protected by copyright. These rights will need to be identified and appropriate rights management and rights clearance strategies employed to deal with them (see below). Rights registers should be kept by projects, and lodged in repositories to accompany all relevant research outputs.

15. It will be important to give users clear parameters regarding their access to derived data and under what terms and conditions. Options for consideration might include Open Access licences, such as Creative Commons or Science Commons licences (or other similar licences) or repository licences. In making this decision, the licence should be in line with the IPR statement (above) and aims and objectives of the project. Institutions should be aware of the associated risks and benefits associated with such a decision (such as the irrevocability of the licences and also the pros and cons of the various licences that are available). If these deliverables are then licensed for use by users, there is the need for compatibility between the permissions secured from the rights holders of the data, and the permissions that users are allowed to access the deliverables. This will need to be planned ahead of time and also used to inform the selection of the appropriate licence for the delivery of the project outputs to users.

16. Risk assessments need to take into account and suitably address the complexity of the IPR issues within projects dealing with derived data. For example, if third party rights need to be cleared, mitigation strategies will need to be clearly outlined to ensure that risks are suitably calculated and addressed. They might include the investment of resources in unpicking IPR issues and the clearance of rights and/or robust notice and take down procedures and policies.

17. All project partners will need to sign a consortium agreement clarifying roles, responsibilities towards the clearance of IPR and who owns the IPR in any derived data.

18. JISC-funded projects should ensure that whilst their IPR statement will reflect the "**what"** relating to IPR and licensing, the work packages within the project plan should

include a legal and licensing component to reflect “**how”** the project will address licensing and IPR issues in derived data (as below). This should ideally cover:

- The key legal issues
- A clear differentiation between the third party rights which need clearing, the IPR issues arising in the deliverables and the licensing implications of negotiating rights with third parties and how the outputs will be made available to users
- The responsibilities for dealing with these issues
- The allocated resources in terms of staffing and administration costs (as well as suitable financial considerations reflected in the Budget)
- Timeframes
- Rights registers.

19. JISC-funded projects may need to develop strategies for the clearance of third party rights regarding derived data. For example, there may be the incorporation of a number of other inputs and experiences of other projects, which will need to be identified as far as possible within the project plan. In this respect, it will be important to clarify:
    - The types of rights that need to be cleared, such as copyright and database rights
    - Establishing processes to ensure that all rights have been cleared to use these third party generated inputs
    - Drafting suitable agreements with third party contractors to ensure that they assign their rights
    - Developing terms and conditions governing user-generated material and material submitted by users, which should also be addressed to ensure that material that is submitted is legal, does not infringe third party rights etc. It might also be important to clarify who owns the rights in the material that is submitted, and/or material that is generated as a result of the submission and what can be done with it by the project partners and/or the user. Other issues such as disclaimers should be put in place to cover legal issues.
    - JISC should coordinate with other ‘research’ funders to ensure a consistent approach towards IPR issues in derived data, and should ensure that it adopts a consistent IPR approach in all its development activities that involve derived data.

We would like to thank John McNaught, Associate Director, National Centre for Text Mining, Lorraine Estelle, CEO, JISC Collections, Neil Jacobs, JISC, Alison Turner, JISC and David Medyckyj-Scott, EDINA for their assistance.