Challenges and Opportunities related to data portability in the context of open science

Andrew Katz
Visiting Researcher, University of Skövde
Partner, Moorcrofts LLP, UK

MOORCROFTS
CORPORATE LAW

Context

MOORCROFTS
CORPORATE LAW

More data was recorded in the last two years than in the entire history of the human race.

MOORCROFTS
CORPORATE LAW

By 2020, 1.4Mb of data will be recorded every second.

MOORCROFTS
CORPORATE LAW

By 2020, 1.4Mb of data will be recorded every second.

For every person on the planet

MOORCROFTS
CORPORATE LAW

By 2020, 44 trillion gigabytes of data will have been recorded.

MOORCROFTS
CORPORATE LAW

Over a billion smartphones are sold every year.

MOORCROFTS
CORPORATE LAW

Less than 0.5% of the data ever collected has been analysed in some way.

Source of this and previous slides: Forbes 2015

MOORCROFTS
CORPORATE LAW

Hot issues

- File formats: obsolescence and patents
- Copyright and database right
  - Why database right ?
  - Text and data mining
  - Computer-generated works
- Personal Data and GDPR

MOORCROFTS
CORPORATE LAW

File Formats - accessing data

MOORCROFTS
CORPORATE LAW

File Formats

Obsolescence:

1. Physical media
2. Filesystem formats
3. File formats
4. Copyright in software to read them
5. Patent issues in the format/ software

MOORCROFTS
CORPORATE LAW

File Formats - Copyright

- Can we get a licence to the software?
- Is any compatible software available?
- Open source is better than proprietary (longevity)
- Reverse engineering of file formats is possible, as is reverse engineering software for interoperability purposes.

(Computer Programs Directive 2009/24/EC)

MOORCROFTS
CORPORATE LAW

File Formats - Patents

- Even open standards can be problematic from a standards perspective
- Patents ***do*** expire, but there can be a >20 year period when they are potentially a problem
- Difficult to legally implement software standards which may be covered by patents

MOORCROFTS
CORPORATE LAW

SSOs and Patents

- ISO, IEC, ITU-T, ITU-R Common Policy
- Organisations involved in standards setting must declare patents
- SEP: Standard-essential patents
- Declaration form: organisations declare whether they will offer to license their patent(s) on
  - RAND terms with no royalty
  - RAND terms
  - No licence at all.
- Forms entered into SSO's database against the relevant standard

MOORCROFTS
CORPORATE LAW

Issues with SSOs

- Declarants slow to respond
- Declarants do not necessarily disclose specific patents
- Likely that declarants are still registered on the database after the relevant patents have expired
- Do not know the extent to which the patent necessarily impinges on a specific implementation (if at all)
- One declarant refused to provide a licence which is compatible with GPLv3
- One declarant tried to refer licensing questions back to ISO

MOORCROFTS
CORPORATE LAW

Standards Setting Orgs

- Patent clearance is difficult
- Open source licences: RAND-Z incompatibilities
- SSO databases are:
  - Not comprehensive
  - Not up to date
  - Only cover entities which are involved in the standards setting process
  - Do not (always) refer to specific patents
  - Only indicate in vague terms willingness (or otherwise) to grant licences

‘On Implementation of Open Standards in Software’, Lundell, Gamalielsson, Katz 2015 http://www.igi-global.com/article/on-implementation-of-open-standards-in-software/148742

MOORCROFTS
CORPORATE LAW

Example - PDF/A-2

- PDF/A-2 (ISO 19005-2:2011)
- contains normative references to other standards (maintained by ISO and other SDOs)
- inherent parts of the ISO 19005-2:2011 standard.
- e.g: Part 2 of the ISO standard - JPEG 2000
- Several patent declarations can be identified in the ISO patent database for JPEG 2000.
- The same declarations cannot be found when searching for declarations made for ISO 19005-2:2011 (PDF/A-2 itself)
- necessary to manually search the ISO patent database for all normative references at all levels).

Lundell, Gamalielsson, Katz 2015

MOORCROFTS
CORPORATE LAW

JPEG 2000 ISO/IEC 15444

- 13 parts and 45 standard documents (total cost 2718 CHF - about 24,000SEK)
- 16 organisations declared SEPs in ISO database - contacted by email and post
- Only 3 responded (after reminders)
    - One unwilling to grant a license for their patents that would allow implementation in software provided under common open source licenses
    - One response declined to clarify which patents they control (the response was “we have at least 3 patents”)
    - One response explicitly stated that they decline to respond
- Contact details out-of-date for 5 (of 16) organisations

Lundell, Gamalielsson, Katz 2015

MOORCROFTS
CORPORATE LAW

File Formats - Archive Formats

- Open standards?
- Well-understood and mature
- Availability of software
- ‘Locked in’ formats
- Lossy/lossless?

MOORCROFTS
CORPORATE LAW

Ongoing research (LIM-IT)

- Investigations focused on trying to obtain all necessary rights for implementation of (a selected set) of commonly used file formats
- Development of recommendations
  - What can (and should) a company do?
  - What can (and should) a public sector organisation do?
  - What can (and should) policy makers (EU) do?

MOORCROFTS
CORPORATE LAW

Copyright and Database Right

Open Licences and Text and Data Mining

MOORCROFTS
CORPORATE LAW

Open Content Licensing

- Creative Commons
  - ‘no derivatives’
  - attribution
  - share alike
  - non-commercial

MOORCROFTS
CORPORATE LAW

No Derivatives

- Difficult to quote (rely on fair use/fair dealing)
- Difficult to extract data (although facts aren't *supposed* to be covered by copyright)
- Can't translate into other languages
- Difficult to data mine.

MOORCROFTS
CORPORATE LAW

Attribution

- Cumbersome when data is extracted

MOORCROFTS
CORPORATE LAW

Share-alike

- potential problems with licence compatibility
- not appropriate for data

MOORCROFTS
CORPORATE LAW

Non-commercial

- Difficult to understand what 'commercial' means.
- Does not meet open source/open content definition.

MOORCROFTS
CORPORATE LAW

Database Right

- Is it necessary at all?
- US manages just fine without it
- 2017 consultation from the EU
  - I represent 9% of the UK respondents (1/11)
  - 30 individuals responded in the whole of the EU:
  - 113 organisations

MOORCROFTS
CORPORATE LAW

“Introduced to stimulate the production of databases in Europe, the new instrument has had no proven impact on the production of databases.”

http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf

MOORCROFTS
CORPORATE LAW

Text and Data Mining

- ‘Mining’ is a bad analogy: implies digging for a nugget
- ‘Mapping’ or ‘analysing’ would be better
- Concerned with extracting facts ***in*** and ***about*** copyright materials
- Facts are not supposed to be copyrightable, so should not be an issue
- ‘This Act does not apply to …facts and data’ (Copyright Act 1997 §2.7, Estonia)
- But…you need to have a copy of the material in the first place

MOORCROFTS
CORPORATE LAW

TDM Exemptions

- EU: Permits “Research Organisations” to undertake TDM
  - need lawful access
  - Research organisations equal non-profits, Universities etc. Public-private partnerships may be excluded, if there’s business has decisive control
- UK: 29A and Schedule 2(2)1D of the Copyright Designs and Patents Act 1988:
  - allows TDM for non-commercial research
  - interpreted very restrictively
  - Need lawful access
- US: “Google Books” - fair use exception
- Japan - do not need licence to the work in question, but excludes databases

MOORCROFTS
CORPORATE LAW

TDM - Solution?

- Drafting legislation for one small jurisdiction
- ‘Safe harbour’ based on TDM activities being ring fenced within a specific facility, subject to security,
- the works may only be used for TDM and:
- the output must **not** contain any derivative work (in the legal sense) of any of the input documents (unless licensed).

MOORCROFTS
CORPORATE LAW

GDPR Research Exemption

- Covers personal data only
- Art 9.2.j
- Allows the data to be processed for reasons other than the purpose they were collected for provided that the output does not contain personal data, or decisions are taken based on personal data
- Subject to discretion of member states on implementation.

MOORCROFTS
CORPORATE LAW

GDPR Research Exemption

- Data must be collected lawfully in the first place
- We advise establishing a ‘research division’ with its own server, security, policies, training, and an internal prohibition on data flowing from research>operations
- Concerns about triangulation.

MOORCROFTS
CORPORATE LAW

Machine Generated Works

- UK allows works generated by machine (e.g. computer) to be protected by the person who ‘made the arrangements’ to create the work - e.g. who used a random number generator to write music.

- Does this mean that the author of AI software can claim to have ‘made the arrangements’ for all of the output? Or the person making use of the software?

- Researchers have identified significant problems and we are launching a research group later this year to consider.

MOORCROFTS
CORPORATE LAW

Summary

- Data is becoming more prevalent and more accessible.
- There is tension between those collecting and those wanting to use.
- In some cases, the legal context has unintended consequences.
- Although the issues are starting to become recognised, practice and legislation (as ever) will take time to catch up with technology.

MOORCROFTS
CORPORATE LAW