



University  
of Glasgow



# The Law of Data Scraping: A review of UK law on text and data mining

Dr Sheona Burrow, 2021

**WORLD  
CHANGING  
GLASGOW**



## Scope of Paper

- Review of UK law on text and data mining (TDM), web/data scraping
- Map the interpretation of the copyright exceptions potentially applicable to TDM



## Legal Regimes – Property based

- Copyright law – protects (creative) ‘works’ - harmonized range of exceptions, including TDM exception (for ‘*research for a non-commercial purpose*’) since 2014 – reproduction/adaptation rights for owners
- Sui Generis Database Right (SGDR) – protects information that has been ‘obtained, verified or presented’ in a database (but not created!) from unauthorized extraction and reutilization – exception for purpose of ‘*illustration for teaching or research and not for any commercial purpose*’



## Legal Regimes – Liability Based

- Data Protection Law – data processing must comply with the detailed regime set out in GDPR
- Contract law – commercial/non-commercial/research – some exceptions cannot be overridden by contract
- Confidentiality – common law – obligation of confidence
- Competition law



## Some of the problems

- Information is usually ‘low IP’
- A (temporary) copy or extraction is required for TDM – some level of adaptation or reutilization may happen.
- Is the information created or obtained?
- Exceptions cannot be contracted out of, but if copyright and SGDR don’t apply does contract law have free rein?
- Single source databases = anti-competitive
- Interpretation of terms – ‘research’, ‘non-commercial’, ‘not for commercial purposes’ and ‘commercial’ – anticipation of future use?



## Scenarios

- Sports databases & betting – limited protection except confidentiality – deemed to be ‘created’ rather than obtained – match data – created – not protected – live data is obtained
- Price Databases – if no SGDR (created/obtained?) then t’s and c’s can apply
- Location information – personal data protected
- Compilations of works - unclear



# Methodology

- Search of legal databases, institutional databases and ‘grey’ literature using key terms\*
- Seeking literature on what is permissible/prohibited in relation to data use, excluded:
  - ‘How to’ conduct TDM and other data science methodologies;
  - Promotion/celebration of TDM;

\*‘data mining’, ‘text mining’, ‘text and data mining’, ‘TDM’, ‘web scraping’, ‘data scraping’, ‘text scraping’, ‘copyright exceptions’, ‘data’



## Results

- Most popular terms: ‘data mining’ and ‘web scraping.’
- Case law is limited and not useful for those seeking to engage in TDM.
- Academic literature limited and repetitive.
- No case law yet on new TDM exception.





## Problems for TDM users?

- Variety of overlapping legal protection to consider
- Exceptions are variable by legal regime
- Fact dependent case law
- When not protected by copyright/SGDR then contract may override – but often cannot be drafted appropriately
- Open licensing contractual terms may not be clear (“non-commercial”)



University  
of Glasgow



# What's the problem? Some cases...



**(1) *British Horseracing Board Ltd v William Hill Organisation Ltd* [2001] EWCA Civ 1268**  
Database of horse racing fixtures

Copyright



SGDR





## **(2) *Ryanair Ltd v PR Aviation BV* (C-30/14)**

**EU:C:2015:10; [2015] 2 All E.R. (Comm) 455**

Website database of flights available, times, prices

Copyright



Contract



SGDR





**(3) *Racing Partnership Ltd v Done Brothers (Cash Betting) Ltd* [2019] EWHC 1156 (Ch); [2020] Ch. 289; [2019] 5 WLUK 112 (Ch D)**  
Horsereading data collected live at racecourses

Copyright



Contract



SGDR



Confidentiality?





## **(4) *77M Ltd v Ordnance Survey Ltd* [2019] EWHC 3007 (Ch)**

Database of geo-spatial co-ordinates of all addresses in the UK created from multiple databases

Copyright



Contract



SGDR?



Confidentiality





University  
of Glasgow



# Thank you

[Sheona.Burrow@glasgow.ac.uk](mailto:Sheona.Burrow@glasgow.ac.uk)

[sheonaburrow@gmail.com](mailto:sheonaburrow@gmail.com)

#UofGWorldChangers



@UofGlasgow