



LEGAL ASPECTS OF CONTENT MINING

copyright
publisher licences
country-specific legislation



MINING REQUIRES COPYING



- PDFs are not easily machine-readable
- Need to **copy** and adapt into digital format specific to particular content required

COPYRIGHT

Gives the **owner** rights to authorise (or not) any of the “restricted acts”



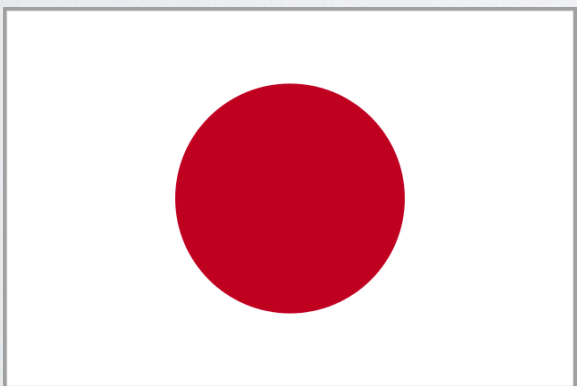
- copying
- adapting
- disseminating all or “substantial” part
 - not just “most of”
 - substantial = important



COPYRIGHT EXCEPTIONS



Doing such “restricted” acts results in infringement -
can be sued for this



However there are exceptions where copying is
allowed without permission or paying fees

Only two countries (UK and Japan) have specific TDM exceptions



DATABASES

“a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”



DATABASES

- in rare cases, databases are **not protected**
- **copyright** applies if the contents constitute “author’s own intellectual creation”
- **database right** only achieved if “substantial investment in obtaining, verifying or presenting the contents” *i.e.* not automatic for all databases



DATABASE RIGHT

- prevents the extraction/re-utilisation of whole or substantial part, qualitatively or quantitatively
- protection for 15 years - renewable if substantially transformed



MINING DATABASES

Lawful use (regardless of contractual terms to the contrary):

- access and use the database
- extract and re-utilise insubstantial parts
- with acknowledgement, extract substantial parts without re-utilising for non-commercial teaching or research



TDM COPYRIGHT EXCEPTION

- active in the UK since June 2014
- anyone who has lawful access to the work can make copies for the purpose of computational analysis
- must be for **non-commercial** research
- “contract terms that stop researchers making copies to carry out text and data mining will be unenforceable”



TDM COPYRIGHT EXCEPTION

- does not apply to database right (even if database is also copyright protected)
- UK researchers do not have to ask permission/pay fees
- defining commercial/non-commercial is very grey area



PUBLISHERS AND MINING

- Have historically charged high fees and placed restrictions
 - must use their own API
 - slow/no response to TDM requests
- Researcher must approach many publishers for permission
- These barriers have slowed TDM technology advances
- So TDM experiments are often limited to OA materials

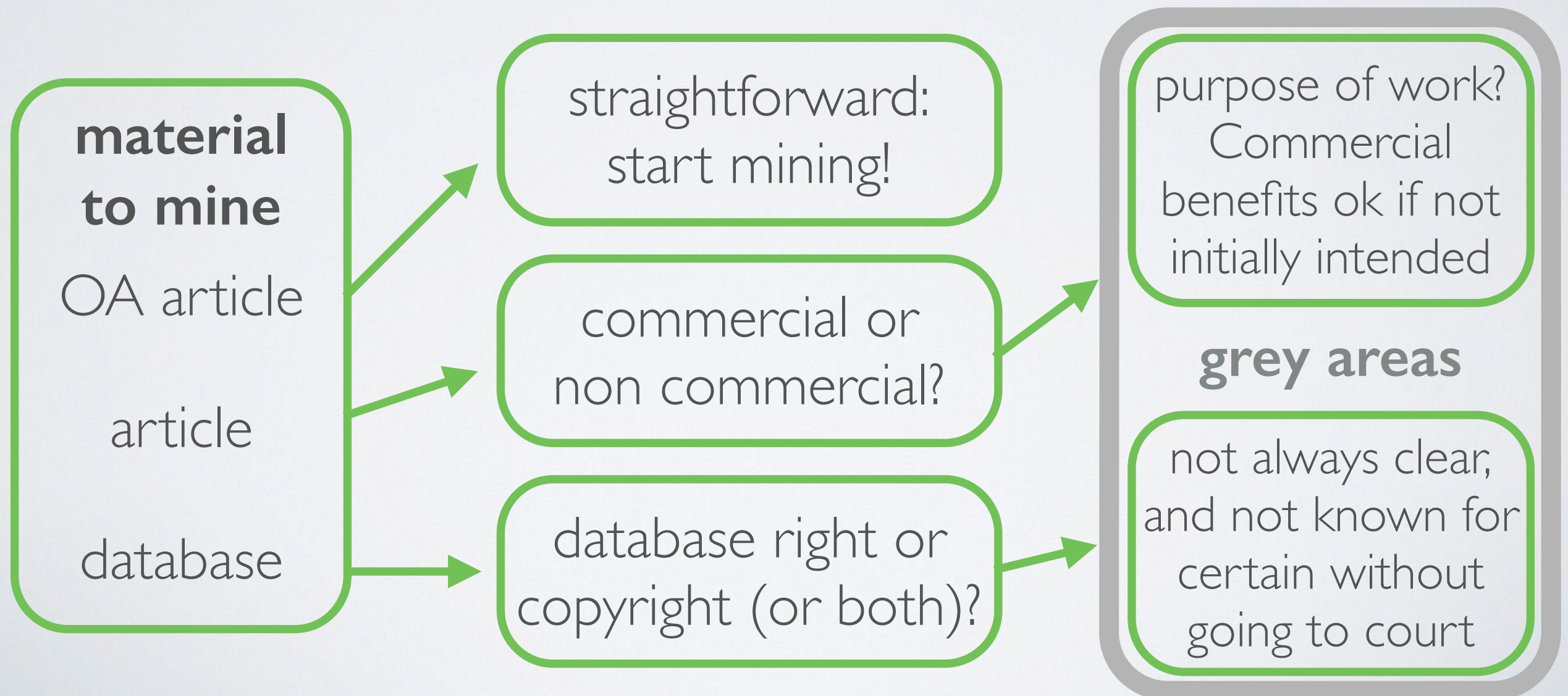


PUBLISHERS AND MINING

- Now offering TDM licences and policies
- Often imposing unfair and **unenforceable** constraints
- Need a test case that goes to court so things become clearer



TDM LEGAL WORKFLOW





CONCLUSION

- new copyright exception is great news for TDM
- law generally is unclear due to lack of case studies
- currently it is about **risk management** and the importance of the work and researcher reputation