

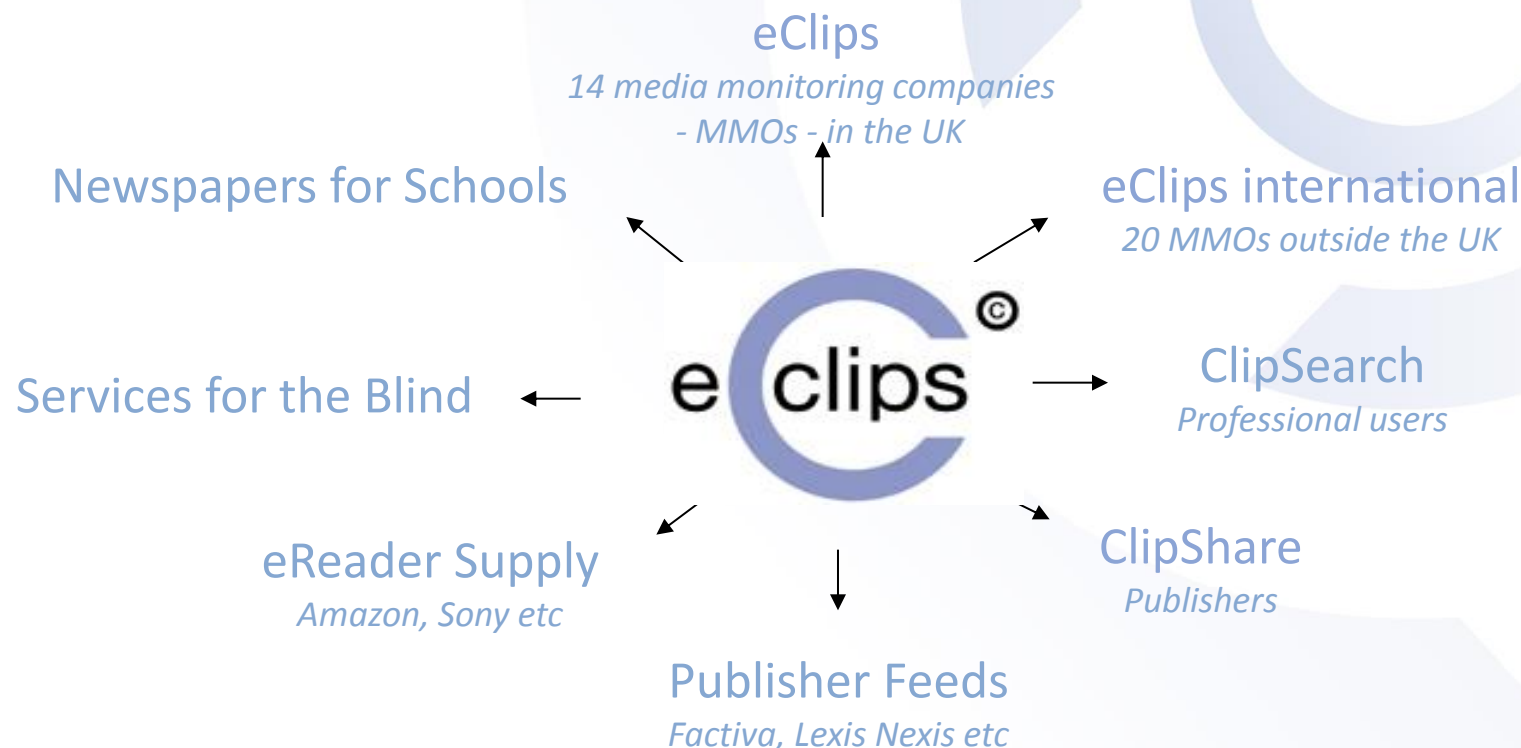


newspaper licensing agency  
Respect for Copyright

# Text Mining and Newspaper Content

Andrew Hughes  
ahughes@nla.co.uk

# NLA eClips – one database, 500 newspapers, many services



- NLA licenses UK newspaper use, by businesses and intermediaries
- NLA creates *eClips*, a 500+ newspaper database

# Text mining is copying

## Text mining requires copying of all data

- To establish data patterns and connections, computers need to index data
- Every word on every page has to be copied
- Once the copy exists, it needs to be managed

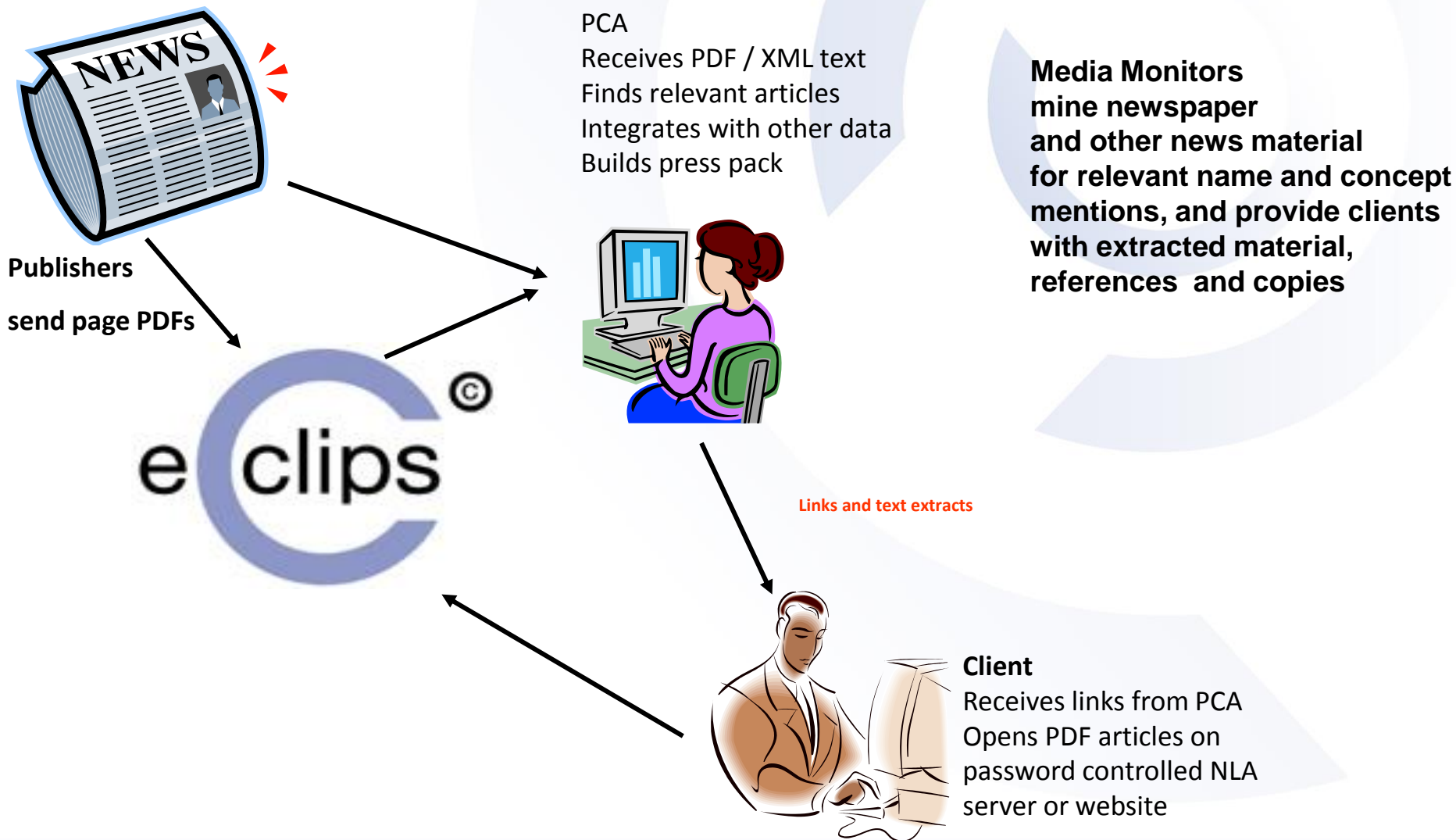
## Copying requires access to data:

- Indexing can only happen on either;-
  - Publisher databases
    - risk of damage and disruption unless managed, and expense
  - Copy provided to text miners' database
    - Costs, and control risks for publishers

## Licence and physical controls needed to prevent commercial damage

- NLA eClips database generates £5-7m value pa
- NLA experience is miners (in NLA case, media monitoring businesses) have a very poor compliance record

# Press packs are produced by text mining



## Mining can be lucrative – the Meltwater case

- 2002 – 2012 copied (mined) entire contents of 1,000 newspaper websites daily, without permission, refusing notice to stop
  - Sold links, text and text extracts to users
  - Refused an NLA licence allowing this activity
  - Competed with many NLA licensed suppliers of press cuttings (MMOs)
  - Claimed it was 'only indexing' not copying
  - Claimed (trivial amounts) traffic to publishers was a benefit
- Meltwater lost the legal case 2011
  - Indexing is copying, headlines are copies, text extracts are copies
- Meltwater lost the Copyright Tribunal case 2012
  - They should pay comparable fees to licensed MMOs
- Meltwater UK turnover £10m
- Contribution to UK creators
  - Pre court action est < £30k
  - Post court action est >£1m



**What public interest is served by exempting this activity?**

## Publisher licensing works well

Meltwater are now thriving within the NLA rules

- Now licensed by NLA, and users licensed
- Operating on fair and equal terms with competitors
- Modest revenue flowing to publishers and suppliers
- Economic alignment with publishers created
- Exploring eClips solutions to improve services



- Leading global online media monitoring solution
- Monitor over 162,000 online publications
- Analysis, Newsfeed, and Newsletter capabilities
- Intuitive and easy-to-use platform
- Easily scaled and tailored to your organization
- Unlimited consultation from Meltwater for lifetime of subscription

**What public interest is served by exempting this commercial business from copyright?**

# Text mining must be managed by publishers

- There are substantial risks and costs for publishers in text mining

## **Risks**

- Technical errors by miners
- Data integrity
- Commercial malpractice

## **Costs**

- Technical loads on systems
- Management of copies and use
- Opportunity costs

- Any proposals should be;-

- based on fully researched and proven benefits
- focused carefully and narrowly on specific applications
- recognise potential commercial free riding
- accommodate and allow for technical risks + expense

