USER GUIDE

COPYRIGHT EVIDENCE WIKI

INTRODUCTION

Welcome to the Copyright Evidence Wiki User Guide!

Copyright Evidence is a digital resource developed by the CREATe Centre at the University of Glasgow. The aim of the Wiki is to construct a complete catalogue of existing empirical evidence relevant to copyright policy in order to inform public debate. The evidence is coded by many categories, including country, industry, funder and research method, offering an in-depth view of existing findings.

This guide provides a user-friendly introduction on how to navigate and search the Wiki, build visualisations and understand the key terms we use to code studies.

CONTENTS

copyright **EVIDENCE**

0 2

HOW TO USE THE WIKI

The Wiki is intended to be used as a dynamic, interactive literature review tool. The "**Semantic Drilldown**" feature can be used to design a tailored search, allowing you to filter according to a range of factors, including country, industry and research method. The Wiki is also fully text-searchable, meaning you can look for specific terms that may not be captured by these factors (e.g. "Facebook", "newspaper").

After you have tailored your search, you may want to consider using the "**Visualisation**" feature to represent your findings. The Visualisation feature contains a variety of graphs and mapping features:

Bar Charts: We use a number of bar charts to represent categorical data, including e.g. Industry and Year of Study. These charts show rectangular bars with heights or lengths proportional to their values and may be sub-divided by colour codes to represent the prevalence of a certain category.

Geolocation: The Geolocation heat map shows the density or magnitude of studies focussing on a particular country. The darker the colour (blue), the more studies from that particular country, whereas the paler the colour (cream) the less. For example, the dark blue colour of Australia indicates that many studies focus on this country.

copyright
EVIDENCE

03

Heat Map Industry/Policy Issue table: The Industry/Policy Issue heat map shows the density or magnitude of studies based in particular **Industries**, across particular **Policy Issues**. The x (horizontal) axis represents industries, the y (vertical) axis represents policy issues (defined in more detail in the Glossary). The darker the colour (blue), the more studies from the cross-section of that Industry/Policy Issue, whereas the paler the colour (cream) the less. For example, this heat map may demonstrate that there are many studies on Enforcement issues in the music Industry.

Pie/Circle Charts: We use pie/circle charts to visualise **Methods** and **Countries**. This circular graphic is divided into slices to illustrate the numerical proportion of each category; the bigger the slice, the bigger the proportion.

Study Network: The study network shows the most frequently cited **Key Related Studies** and how these reference each other. Each node (circle) represents a study, which is connected by lines to demonstrate the connection between it and other studies. Hovering over each node will reveal the names of the related studies.

Word Cloud: The word cloud is a collection or cluster of the most frequently used words in the studies on the Wiki. The bigger and bolder the word appears, the more often it is mentioned, whereas the smaller and paler the word appears, the less frequently. The word cloud can be filtered to show the most frequently used words in studies about e.g. the music industry.

All visualisations can be downloaded as an image using the button.

copyright
EVIDENCE

0 4

GLOSSARY

This section provides an overview and definition of the key terms used on the Wiki to help understand how we capture information.

Abstract: The verbatim copy of the abstract from the study. Where an abstract is not provided, a brief summary in the coder's own words is used instead.

Authors: Authors of the study, entered individually where there are multiple authors. You can click on the author name to see whether they have contributed any other studies to the Wiki.

Comparative study: Where a study compares e.g. countries or industries. We only use this to indicate substantively comparative studies, and it is not used to identify e.g. a brief comparison made in the context of a larger article.

Cross-country study: Where a study covers more than one country.

Country: The relevant country or countries where the study takes place. You can click on the country name to see whether there are any other studies on the Wiki relevant to that country.

Definitive Link(s) to paper and Open Access Link(s) to paper: Definitive Links to the study usually link to e.g. the journal article home page, but may be paywalled. Open Access Links are provided where possible (commonly e.g. SSRN, Zenodo).

Full Citation: The full citation of the study is provided APA-style (e.g. Contributors. (Date). Title. Publication Information.).

copyright
EVIDENCE

05

Fundamental Issues: The Wiki is divided into 5 key theoretical propositions about the copyright incentive, which are represented short-hand with numeric values (1-5). This section is under development, but includes:

- **1 Relationship between protection and economic performance.** This field includes studies that examine the connection between copyright protection (e.g. subject matter, term, scope) and economic performance (e.g. supply, economic growth, welfare). Papers in this category often link different legal and institutional settings to economic performance (e.g. through historical counterfactuals) and also may examine non-IP markets (e.g. recipes, jokes, formats, fashion).

- **2 Relationship between creative process, incentives and legal rules.** This field includes studies that focus on what motivates creators (e.g. attribution, control, remuneration, time allocation, intrinsic and extrinsic motivation). Among others, the papers included under this category connect to the literature on labour markets and earnings, and the production of culture literature (e.g. linking rules on adaptation, sampling, co-authorship to aesthetic outcomes).

- **3 Contracts, harmony and conflict of interests between creators and investors.** This field includes studies that examine the common assumption of a harmony of interests between creators (e.g. authors, performers) and investors (e.g. publishers, producers), a simplifying hypothesis that facilitates analytical solutions, which however finds weak empirical support. Papers included under this category also examine collecting societies and relate to the area of contract theory.

copyright **EVIDENCE**

06

- **4 Effects of protection on industry structure.** This field includes studies that examine the connection between copyright protection, competition and industry structure. Among others, the papers included under this category focus on monopolies, oligopolies, the economics of superstars, new business models, technology adoption and relate to the fields of industrial organisation and competition law.

- **5 Understanding consumption and use.** This field includes studies that examine human behaviour and in particular consumption and use. Among others, the papers included under this category focus on the determinants of unlawful behaviour and changing forms of consumption and use (e.g. user-generated content, social media, streaming) and relate to the areas of behavioural economics and consumer theory.

Funder: Details of any funders/commissioners where possible. This section is mainly included for transparency.

Government/policy report: Whether the study is a government or policy report, usually identified by official markings or sources (e.g. IPO, European Commission) or e.g. a statement that the report has been commissioned by a governmental body.

Industry: The list of industries on the Wiki are based on the categorisation of creative industries as developed by the UK Department for Digital, Culture, Media & Sport (DCMS). For each study, we indicate the industries represented therein (e.g. music, film) as closely as possible.

copyright **EVIDENCE**

07

JEL Codes: JEL Classification Codes originate from the Journal of Economic Literature and are a standard method of classifying scholarly literature. There are 20 primary JEL categories (represented with letters) with numerous subcategories (represented with numbers). We use this where possible to improve our search functionality, though many studies will not explicitly state their JEL code.

Key Related Studies: These are other, key studies which are directly referenced in the text of the study on the Wiki. We give primacy to studies that are referenced more prominently or frequently in the text to infer a direct relationship. Not every Key Related Study will in itself be coded on the Wiki, as many studies rely on e.g. studies on economics which, whilst central to the study itself, is not within the ambit of the Wiki.

Literature review: Usually studies of this kind are a summary of previous studies on a given topic which is synthesised to critically analyse e.g. gaps in current knowledge or limitations of previous studies. This is only used where the study is wholly a literature review, and not to identify e.g. the literature review section of a research article.

Main results of study: Coder's provide a summary, in their own words, of the study's main results. This may include e.g. statistical findings (e.g. "90% of authors think…") or conclusions (e.g. "the study finds three emerging themes…").

Policy Implications: Many studies on the Wiki may have suggestions about how to improve copyright policy based on their findings. The Wiki only includes policy implications from a study where this is explicitly mentioned by the author, or where the suggestions made by the author could be reasonably interpreted and translated into policy instructions.

copyright **EVIDENCE**

0 8

Policy Issues: As the Wiki aims to capture studies which can inform policy and public debate, the Policy Issues category defines 6 key areas, which are represented in short-hand with alphabetical values (A-F):

- **A. Nature and scope of exclusive rights.** This field includes papers that examine policy issues related to the types of works that are eligible for copyright protection and the extent of the protection offered by exclusive rights and moral rights. Among others, the papers included under this category focus on the originality threshold, derivative works, hyperlinking, news aggregation, resale and community norms (including negative space).

- **B. Exceptions.** This field includes papers that examine policy issues related to whether materials which otherwise are subject to exclusive copyright protection should be available for justifiable use without seeking permission and whether existing exceptions and limitations facilitate creative and scientific progress. Among others, the papers included under this category distinguish exceptions and limitations for the purposes of innovation or public policy, open-ended provisions from closed lists, and commercial and non-commercial uses.

- **C. Mass digitisation / orphan works.** This field includes papers that examine policy issues related to the process that enable mass digitisation of copyright protected content. Among others, the papers included under this category focus on potential solutions for orphan works and non-use of cultural works, including licensing schemes and extended collective licensing, and the application of copyright in cultural heritage institutions.

copyright **EVIDENCE**

0 9

- **D. Licensing and business models.** This field includes papers that examine policy issues related to strategies and licensing solutions in the exploitation of copyright protected materials, and how legal markets attempt to match production to consumption. Among others, the papers included under this category examine collecting societies, metadata, copyright exchanges and hubs, windowing, crossborder access, open access/open science and end-user licensing.

- **E. Fair remuneration.** This field includes papers that examine policy issues related to creators' or rightholders' earnings. Among others, the papers included under this category focus on the sources of artistic income, royalty flows, contracts, levies and sales displacement.

- **F. Enforcement.** This field includes papers that examine policy issues related to the optimal way to enforce the private right of copyright. Among others, the papers included under this category focus on quantifying infringement, motivations for infringement, technological measures of protection, intermediary liability, graduated responses, notice and takedowns, criminal sanctions, litigation and court data.

Year of Study: The year in which the study was published.

copyright
EVIDENCE

10

DATA GLOSSARY

Data type: Primary data is data that is collected by researchers directly from the main source (e.g. via participants to a survey or interview). **Secondary** data is data that has already been collected elsewhere and which is made available for other researchers to use (e.g. a pre-existing database detailing income).

Dataset: The dataset section identifies the **Unit** under analysis (e.g. users, companies, countries), the **Sample** size (e.g. 500) and the **Time Period** under which the data was collected (e.g. 2010 – 2012).

Description of Data: A brief overview, in the coder's own words, about how the data was gathered and how it was analysed. This may include information about the number of participants in a study, how they were sourced, which software was used to analyse the data etc. This is usually supplemented by other information in the 'Data' section.

Secondary Data Sources: If possible, and where appropriate, we identify the secondary data source, including names or identifiers (e.g. IMDB, Omeba).

Time period of data collection: The time period that the data was gathered (e.g. 2010 – 2012).

RESEARCH METHODS GLOSSARY

What is empirical research? Empirical research is based on verifiable observation or experience rather than theory or logic. The methods coded on the Wiki refer to how this empirical research was carried out, and how the resulting data was analysed. This section offers an overview of the methods featured in the Wiki, with a brief explanation of each.

QUANTITATIVE DATA COLLECTION METHODS

Studies under this category gather information about quantities and use numerical data.

Experimental (Laboratory, Field, Natural): All types of experiments concern an independent variable (IV) which is manipulated to measure the effects this has on the dependent variable (DV) (e.g. manipulate the amount of money that is given to someone (IV) to see if they are more productive (DV)). In **Laboratory** experiments, the IV is deliberately manipulated in a controlled environment. In a **Field** experiment, the experiment takes place in a natural setting with less control, though the researcher is still able to manipulate the IV. In a **Natural** experiment there is no deliberate manipulation of the IV as it is already naturally changing (e.g. if a website is shutting down in a particular country, this could be the IV which is already naturally changing).

Longitudinal Study: This method studies a subject through observation and repeated re-testing over a period of time.

Quantitative data/text mining: This method usually uses artificial intelligence (AI) technology to derive patterns and trends in the underlying data. Typically, this either deals with structured data (e.g. from a database) or unstructured textual data (e.g. social media feeds).

Snowball sampling: This method of sampling relies on research participants to recruit other participants for a study and is typically used where potential participants are hard to find (perhaps e.g. because of illegal activity). The name derives from the idea that once the snowball is rolling it picks up more snow along the way and becomes larger.

Survey Research: This method uses a survey or questionnaire to gather numerical data (e.g. about income). Surveys generally involve a large audience to collect a large amount of data using a set of predetermined questions.

Web analytic: This method concerns the collection, reporting and analysis of website data (e.g. web traffic).

QUALITATIVE DATA COLLECTION METHODS

Studies under this category gather information through non-numerical data which can be observed and reported but not necessarily measured. These studies typically try to find meanings, opinions, or underlying reasons from its participants.

Archival Research: This method involves investigating the primary sources held in archives, which may include manuscripts, documents, records, objects, sound and audiovisual materials.

Card Sorting: This method involves participants organising topics (cards) into categories in a way that makes sense to them.

Case Study: This method concerns the study of a particular case of e.g. a person, a company, a group, and usually involves various other methods of data collection (e.g. interviews or observation). This method tends to have a narrow focus and provides detailed descriptive data.

Document Research: This method involves an analysis of documents containing information about a scenario or event (e.g. public or personal documents, newspapers etc.)

Ethnography: This method involves observing and interacting with participants in their real-life environment to understand their individual culture (e.g. social interactions, behaviours). This method may also combine with other methods such as e.g. participant observation or interviews.

Focus Groups: This method typically involves a deliberately selected small group of people who are led in a discussion facilitated with a moderator. This is commonly used to gather feedback or discuss a concept.

Historical Methods: This method is typically used by historians to research and write histories of the past, usually by analysing primary sources.

Interview (Structured, Semi-Structured and Unstructured): Interviews can be carried out in person (face-to-face) or via e.g. telephone or Skype. If the interview is **Structured**, the researcher will only ask specific pre-determined questions. If the interview is **Semi-Structured** the researcher may have pre-determined questions which can be elaborated on more freely depending on how the interview develops. If the interview is **Unstructured** no questions are prepared in advance.

Life History: This method involves documenting a personal account of a person's life history over a period of time. This may involve interviewing, examining letters and diaries or archival research.

Participant Observation: This method involves the researcher becoming (or may already be) part of a group that is being observed. The researcher observes what the group do and offers explanations for this.

Qualitative content/text mining: This method usually uses artificial intelligence (AI) technology to derive patterns and trends in the underlying data. In qualitative content/text mining, this AI may be used to capture e.g. feelings, emotions or perception based on patterns of words or phrases.

copyright
EVIDENCE

Survey Research: This method uses a survey or questionnaire to gather non-numerical data (e.g. of piracy or consumer preferences). Surveys generally involve a large audience to collect a large amount of data using a set of predetermined questions.

Visual Ethnography: This method uses photography, video or film to study and interpret social organisations or cultures.

QUANTITATIVE DATA ANALYSIS METHODS

Calibration: Calibration of a model refers to evaluating its goodness-of-fit or accuracy. The researcher sets desired parameters with known good values to evaluate, which can improve confidence levels in the study's results.

Cluster analysis: This analysis technique is used to classify objects or cases (which in quantitative studies may be numerous) into related, meaningful or useful groups called 'clusters'.

Confirmatory Factor Analysis (CFA): This method allows a researcher to figure out if a relationship between a set of observed variables and their underlying construct exists. This is primarily used for verification and robustness (e.g. does a ten-question survey accurately measure one specific factor, such as socioeconomic status).

Correlation and Association: This analysis method measures the association between two or more variables and provides an inference about the strength of that relationship (e.g. as income increases, so does happiness levels).

Decision Tree Method: A decision tree is a flowchart-like structure beginning with a single 'root' that branches into a number of solutions, offering a choice between several courses of action.

Descriptive statistics: This analysis typically provides a summary of the basic features of the data (e.g. 50% of participants are from Europe).

Factor Analysis: This is a method of reducing a large number of variables into a fewer number of factors. The key concept is that multiple observed variables have similar patterns of responses because they are all associated with a latent (not directly measured) variable. For example, if a survey has similar responses about income, education and occupation, these may all be associated with socioeconomic status (the latent variable).

Meta-Analysis: This method involves combining data from multiple studies to identify a common effect (e.g. is the effect consistent from one study to the next).

Multivariate Statistics: This method concerns the study of multiple variables at the same time (e.g. the relationship between gender, education, income).

copyright
EVIDENCE

Quantitative content analysis: This method involves determining the presence of certain words or data in a text, and may be expressed as a frequency (e.g. "piracy" mentioned 50 times in a text).

Regression Analysis: This analysis method defines the relationship between two or more variables and is used to predict whether one variable influences another (e.g. does a higher income improve a person's willingness to pay for goods).

Structural Equation Modelling: This method involves representing, estimating and testing the network of relationships between variables using a combination of factor analysis and multivariate regression.

Social Network Analysis: This method investigates social structures through the use of networks and graphs. This is graphically represented in terms of individual nodes and the relationships that connect them.

Social Sequence Analysis: This method typically uses longitudinal data to provide descriptive accounts of sequence patterns (e.g. of turn-taking in a conversation). The overall picture of this data can be used to discover characteristics of the sequences.

QUALITATIVE DATA ANALYSIS METHODS

Abduction/Retroduction: Abduction involves making an inference based on an incomplete set of observations to provide the likeliest possible explanation (i.e. the best you can do with the data at hand). **Retroduction** is similar, but the underlying data has some known or assumed rules or observations that include predicates or predictors (e.g. how police can determine initial suspects via means, motive, opportunity).

Discourse Analysis: This method involves the analysis of written or spoken language, and usually explores the meanings produced by this language use and its context.

Ethnographic/narrative analysis: This method involves the textual interpretative of a narrative by a teller, focussing on interpretation, understanding and representation. This is usually done holistically to create a more complete picture of the relevant group, society etc.

Grounded Theory: This method involves developing a theory to explain empirical phenomena which is developed or 'grounded' in actual data. Typically, this involves a systematic inductive method of coding the data to determine larger themes or theories.

Legal Analysis: This method involves the interpretation of law, usually by applying a legal rule to the facts of a situation.

copyright
EVIDENCE

Qualitative Coding / Sorting: This method concerns the process of labelling and organising qualitative data to identify themes and relationships between them. The researcher might assign labels to certain phrases that represent important or recurring themes in each response.

Textual Content Analysis: This method involves analysing the content of a text to describe and interpret the characteristics of that text.

Triangulation: This method is typically used to assure the validity of research, and involves synthesising data gathered from more than one method on the same topic/issue (e.g. using a combination of surveys, interviews and discourse analysis to determine understandings of piracy).

Visual / Other Content Analysis: This method derives meaningful descriptors from image or video data.

copyright
EVIDENCE