Geomorphology 249 (2015) 52–68

Contents lists available at ScienceDirect

Geomorphology

journal homepage: www.elsevier.com/locate/geomorph

Enriching Great Britain's National Landslide Database by searching newspaper archives

Faith E. Taylor [a,*], Bruce D. Malamud [a], Katy Freeborough [b], David Demeritt [a]

[a] Department of Geography, King's College London, Strand, London WC2R 2LS, UK
[b] British Geological Survey, Nicker Hill, Keyworth, Nottinghamshire NG12 5GG, UK

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form 12 May 2015
Accepted 18 May 2015
Available online 27 May 2015

Keywords:
Landslide
Nexis UK
Hazard database
Newspaper
Archive

ABSTRACT

Our understanding of where landslide hazard and impact will be greatest is largely based on our knowledge of past events. Here, we present a method to supplement existing records of landslides in Great Britain by searching an electronic archive of regional newspapers. In Great Britain, the British Geological Survey (BGS) is responsible for updating and maintaining records of landslide events and their impacts in the National Landslide Database (NLD). The NLD contains records of more than 16,500 landslide events in Great Britain. Data sources for the NLD include field surveys, academic articles, grey literature, news, public reports and, since 2012, social media. We aim to supplement the richness of the NLD by (i) identifying additional landslide events, (ii) acting as an additional source of confirmation of events existing in the NLD and (iii) adding more detail to existing database entries. This is done by systematically searching the Nexis UK digital archive of 568 regional newspapers published in the UK. In this paper, we construct a robust Boolean search criterion by experimenting with landslide terminology for four training periods. We then apply this search to all articles published in 2006 and 2012. This resulted in the addition of 111 records of landslide events to the NLD over the 2 years investigated (2006 and 2012). We also find that we were able to obtain information about landslide impact for 60–90% of landslide events identified from newspaper articles. Spatial and temporal patterns of additional landslides identified from newspaper articles are broadly in line with those existing in the NLD, confirming that the NLD is a representative sample of landsliding in Great Britain. This method could now be applied to more time periods and/or other hazards to add richness to databases and thus improve our ability to forecast future events based on records of past events.

1. Introduction

Risk management decisions can only ever be as good as the risk assessments upon which they rest. The United Nations Hyogo Framework for Action on Disaster Risk Reduction (UN, 2005) identifies the development and improvement of relevant databases as a key capacity-building priority. In the particular case of landslide risk, the limitations of existing landslide inventories have been repeatedly highlighted as the greatest source of error in the landslide susceptibility and risk maps used to inform land-use planning and other mitigation measures (van Westen et al., 2006; Fell et al., 2008). Better data are also important for estimating landslide damage functions and thus for assessing risk in the classic sense of the combined probability and consequences of suffering landslide losses (Fuchs et al., 2007; Quan Luna et al., 2011).

In Great Britain, landslides commonly occur due to physical factors such as coastal erosion and maritime climate, particularly during very wet seasons (Jones and Lee, 1994; Bromhead and Ibsen, 2006). Coupled with vulnerability factors such as high population densities and high-value infrastructure, impacts from landslide events range from economic losses and infrastructure damage, disruption, injuries and (less commonly) fatalities (Pennington et al., 2009). For example, in 2012 Great Britain experienced the highest monthly rainfalls for the last hundred years in many regions (Parry et al., 2013). This resulted in approximately five times as many landslides as usually recorded (Pennington and Harrison, 2013), impacts such as major transport disruptions, evacuations and four fatalities (Pennington et al., 2015-in this issue). These losses have peaked policy interest in better understanding landslide impact and in developing a country-wide landslide hazard impact model to forecast and thereby help prevent them in future (Met Office, 2013).

The principal source of data regarding landslide occurrence in Great Britain, what causes them and the history of their impacts is the National Landslide Database of Great Britain (NLD) (described in detail in Section 2.2). The NLD is an archive of the location, date, characteristics and impact of landsliding in the past, with records dating from the last glaciation to present (Foster et al., 2008). First created in the early 1980s by Geomorphological Services Ltd, the NLD is now maintained and constantly updated by the British Geological Survey (BGS) (Foster et al., 2008). Since its creation, the strategies of data collection have been variable, due to shifts in the underlying resources available, change

* Corresponding author.
E-mail address: faith.taylor@kcl.ac.uk (F.E. Taylor).

http://dx.doi.org/10.1016/j.geomorph.2015.05.019

in available technologies and variation in the intended applications of the database (Pennington et al., 2015-in this issue). The variation in the methods and intensity of past data collection make it reasonable to assume that there are additional landslide events to be found, and more information to be added about existing landslides in the NLD.

In this paper, we present a method to increase the richness of the NLD by searching a digital archive of 568 regional newspapers for articles referring to landslide events. Our aim is not to 'complete' the NLD, but rather to complement existing sources by providing more and richer information about landslide phenomena in Great Britain. In particular, we demonstrate the capacity of this method to enrich the NLD in two ways: (i) adding records of additional landslide events not previously documented in the NLD and (ii) supplementing currently recorded NLD landslide event information, particularly about impacts. As this method draws consistently upon an independent dataset, comparing the results to the contents of the NLD can also provide a way to assess potential bias in the NLD and enhance overall confidence in its data. The method we present here could also be applied to enhance understanding of other natural hazards, such as surface water flooding, whose incidence and impacts are not systematically recorded in existing datasets, particularly when examining records pre-remote sensing (Moores and Rees, 2011; Hurford et al., 2012).

This paper is organised as follows: In Section 2, we discuss the broader difficulties of producing landslide inventories and how these relate to the NLD. We then consider the potential of newspaper articles as a supplementary source of landslide inventory data and review existing studies using this approach before introducing the particular newspaper archive used in our research. In Section 3, we describe the methodology we developed for searching and filtering digital archives of regional newspapers to collect news stories about landslide events and extract factual information from them to enrich the NLD. Then in Section 4, we present results of our newspaper searches for two search periods. In Section 5 we discuss the implications and uncertainties of our methodology and how this methodology might be applied in other contexts. In Section 6 we summarise results and draw conclusions.

2. Background

2.1. Landslide inventories and databases

Detailed information about the nature of past events is important for understanding, predicting and managing landslide risk (Guzzetti et al., 2005, 2012). Van Westen et al. (2006) identify four basic types of information about past landsliding needed to support risk assessment and management:

(i) Inventories of landslides
(ii) The environment surrounding the landslide
(iii) What triggered the landslide
(iv) What elements are/were at risk.

Of the four categories given above, van Westen et al. (2008) and Van Den Eeckhaut and Hervás (2012) demonstrate that the first category, landslide inventories, is the most important when considering potential risk for the future.

Compiling such inventories is complicated by a number of factors, including the following: (i) There are first order conceptual questions about the definition of a landslide 'event' to be recorded as distinct from a landslide triggering event (e.g., an earthquake or heavy rainfall) (Kirschbaum et al., 2010). (ii) Compared to other hazards (e.g., earthquakes or extremes of temperatures), where we often have direct instrumental measurements of the phenomena over a wide region (e.g., ground motion, air temperature), landslide deposits (and associated erosional surfaces) observed on the ground are the *outcome* of a set of interacting processes (Guzzetti et al., 1999) that are rarely feasible to measure systematically instrumentally. Consequently, to produce a landslide inventory, one must actively search for them across a landscape, through methods such as remote sensing and photogrammetry (Soeters and van Westen, 1996), field investigations (Brunsden, 1985), public reporting/interviews and archival research (Salvati et al., 2009) or a combination thereof (Guzzetti et al., 2012). (iii) It can also be difficult to identify and extract landslide events from public databases. For example, in the UK the Highways Agency Road Impact Database, landslides do not have a specific event code. Landslides and engineered slope failures are sometimes noted in a free text field but are more commonly recorded in their database of traffic disruption as "other" (Met Office, personal communication, March 2014).

For the above three reasons, it is rare to have databases of all landslides that have occurred over a region within a given time period, and there may be biases towards locations where humans are affected (Carrara et al., 2003) or larger landslides that are more discernible in imagery/field studies (Wills and McCrink, 2002). The 'completeness' of an inventory will also be affected by the time lag between the landslides occurring and when they are inventoried, as smaller landslides may be eroded/erased from the landscape within a few months of occurring (Malamud et al., 2004; Bell et al., 2012). In a survey of 22 European countries that have or are developing national landslide databases, Van Den Eeckhaut and Hervás (2012) found that 68% of respondents estimated the completeness of their country's database to be less than 50%.

The above difficulties with the completeness of landslide inventories limit the quality and predictive power of landslide susceptibility assessment (Galli et al., 2008). Consequently, landslide risk may be under or overestimated depending on the completeness and homogeneity of coverage of the landslide inventory.

2.2. The National Landslide Database (NLD) of Great Britain

The NLD is the most extensive source of information about British landslide occurrence. A metadata description with examples of its content can be found online at BGS (2014a). The NLD currently contains over 16,500 records of individual landslides occurring between the last glaciation and present day. For each landslide, more than 35 possible attributes can be recorded (Foster et al., 2008; Pennington et al., 2015-in this issue). These can broadly be categorised into:

(i) *Landslide location* (Latitude/Longitude and estimation of locational precision)
(ii) *Landslide timing* (date of occurrence or age)
(iii) *Type of landslide* (fixed categories)
(iv) *Cause of landslide* (fixed categories)
(v) *Size of landslide* (free text)
(vi) *Impact of landslide* (number of fatalities, number injured, cost and other free text)
(vii) *Geological setting of landslide* (fixed categories).

Perhaps due to the somewhat episodic nature of landslide activity in Great Britain, policy concern for landsliding has waxed and waned (Gibson et al., 2013), as have resources for NLD data collection and database maintenance, resulting in temporal and spatial variations in database richness. The first national landslide database was initially established in the early 1980s to raise awareness of the nature and distribution of landslides for planning purposes at a local authority level (Foster et al., 2012). As the method employed was a desk-based review of secondary sources such as technical reports, theses, maps and diaries (Jones and Lee, 1994), the spatial extent of records in the original NLD were biased towards locations of human interest, such as high impact landslides or 'classic' field study locations. During the 1990s, sources of revenue from the database were not large enough to fund the maintenance and regular updating of the database and the project was mothballed. In the early 2000s, the Department of the Environment made the database available to the BGS, who over the next few years devoted considerable effort to restructuring, quality controlling, and supplementing this database into a more user-friendly and commercially relevant resource (Foster et al., 2012). As of 2006, the NLD

can be considered to be in its 'contemporary' phase, where information about new landslide events is systematically recorded and added in 'live' (i.e., as and when the BGS hear about a landslide rather than through periodical retrospective studies). In addition to landslides occurring under natural conditions, since 2012 the BGS also records information about failures in engineered slopes, as they often cause considerable human impact (e.g., if a landslide occurs on a railway embankment, this could cause transportation disruption).

Information about landslides is added to the NLD through a number of primary and secondary research channels, which are described in detail in Foster et al. (2012) and Pennington et al. (2015-in this issue). These can broadly be separated into:

- *BGS maps and archive documents* (e.g., field notebooks)
- *BGS field surveys/reports*
- *Academic literature* (books, journal articles, student theses etc.)
- *Aerial photography*
- Searches of *archive media documents* (newspapers)
- Online keyword searches of *current media sources* (newspapers, radio, television, internet)
- *Personal communication* (public, local authorities, land owners, utilities operators)
- Keyword searches of *social media* (Facebook and Twitter) implemented since August 2012
- *Citizen science reporting* via the BGS "report a landslide" web-portal (BGS, 2014b) since 2009 and BGS Twitter profile (@BGSLandslides), implemented in 2012.

From 2008 to 2013, the search of current media which helps inform the NLD, was performed by Meltwater (2014). Meltwater is a subscription media monitoring service aimed primarily at assisting organisations to manage their PR by scanning online media. They provided the BGS with a daily report based on the results returned from an automated Boolean search of a database of 190,000 online sources, including news, social media and blogs (Meltwater, 2014). However, the actual sources searched and how they may have changed over time are commercially confidential.

With the rise of social media, Twitter has become, along with traditional media reports, a primary channel by which the BGS is alerted of landslide events. Where possible, alerts are followed up via field investigation or contact with affected groups/land owners, prior to inclusion in the NLD. Pennington et al. (2015-this issue) estimate that the addition of social media and inclusion of engineered slope failures since 2012, and improved traditional media search strategies, have increased the number of NLD additions per year by a factor of 10 compared to the start of the contemporary phase (2006).

In the following sections, we describe the use of newspaper articles as a source of information about landslide events, introduce the Nexis UK archive of regional newspaper stories and discuss differences between the current media search strategy used by the BGS and that of Nexis UK.

2.3. Newspaper articles as a source of information about hazards

Mass media is generally the first and primary source of information about hazards for the public (Fischer, 1994). Yet, mass media is also used by scientists and practitioners in the field of hazards in a number of ways, with varying levels of depth of engagement with the media:

(i) *First alert.* A news article may be the first way a practitioner hears that a hazard event has happened. From this first alert, s/he may decide whether any follow-up is required (such as a field visit) (e.g., GDACS, 2014; Public Health England, 2014; Pennington et al., 2015-in this issue).

(ii) *Archives (and scientific analysis of archives).* Archives of news stories about various events can be searched to create or add to a database or inventory of hazard occurrence (e.g., Guzzetti et al., 1994; Black and Law, 2004; Llasat et al., 2009; Kirschbaum et al., 2010).

(iii) *Documenting impacts.* Media can be used as a way of documenting impacts of events from desk based studies, both at the time of occurrence and through future updates/press releases and reports (e.g., Tarhule, 2005 for droughts/floods, and Petley et al., 2007 for landslides).

(iv) *Public perception of risk.* Analysis of the interactions between mass media coverage and public understanding of hazards and risk can be performed (Kasperson et al., 1988). For example, media coverage of a particular hazard can be assessed over time to understand changes in how issues such as responsibility are framed (Escobar and Demeritt, 2014) or assessing variation in interest in a particular story over time (Carvalho, 2007).

(v) *Public communication.* Information can be disseminated through interviews and press statements (creation of media) (Peters et al., 2008).

The use of newspaper articles as a proxy for records of various hazards is not a new technique. In a review of proxy records, Trimble (2008) lists examples of studies as early as 1932 using newspaper reports to construct a record of major landslides occurring in Switzerland from AD 1563 onwards (Heim, 1989 [1932]) and in 1946 using newspaper reports to reconstruct a record of flooding in Utah (Woolley et al., 1946). The technique is also well established in historic climate reconstruction (Demeritt, 1991; Brázdil et al., 2005).

Raška et al. (2014) provide an overview of natural hazard databases that use newspaper and other documentary evidence. For landslides, perhaps the most cited national database is the Italian AVI project (available online at CNR-GNDCI, 2014), containing records of >32,000 landside and >29,000 flood events, going back 1000 years, but with most recorded between 1900 and early 2000, of which ~78% of the information comes from newspaper reports (Guzzetti et al., 1994; Guzzetti and Tonelli, 2004). More recently, the growing capacity to search freely available digital archives of global newspaper reports and online sources has prompted the construction of the Durham Fatal Landslide Database, which is a global record of landslides triggered by rainfall that have resulted in fatalities since 2004. For the seven year period, 2004–2010, the database includes 2620 landslides, which resulted in 32,322 fatalities (Petley, 2012). Other examples of landslide databases using newspaper articles as a source of information include Domínguez-Cuesta et al. (1999) in the North of Spain, Glade and Crozier (1996) in New Zealand, Devoli et al. (2007) in Nicaragua and Kirschbaum et al. (2010) at the global scale.

There are clear biases in newspaper articles as a proxy for information about hazards, such as an overemphasis on events with human impact (Carrara et al., 2003), increased media interest following a number of events (Pennington and Harrison, 2013), a focus on high magnitude events or underreporting of low magnitude events (Guzzetti and Tonelli, 2004) and scientific correctness of information (Ibsen and Brunsden, 1996). Nonetheless, the regular publishing intervals (and thus continuous record) (Raška et al., 2014) and relative ease and low associated costs of performing a desk-based study means that analysis of newspaper articles is widely seen as a useful complement to other methods for building hazard databases. For example, in a review by Tschoegl et al. (2006) of 31 major international, regional, national and sub-national hazards databases, newspaper reports are used as a regular and/or major source of records about hazard events in 10 of the databases.

2.4. The Nexis UK regional newspaper archive

In the last decade, there have been considerable advances in the digitisation and indexing of archives of newspapers in the UK, for example, The British Newspaper Archive (British Library, 2014) and The Nineteenth Century Serials Editions (NCSE, 2007). Here we have explored the use of a digital subscription archive, Nexis UK (LexisNexis Academic, 2014), to add richness to the NLD. The archive was chosen due to its national scope, coverage up to present day and the relative ease of searching.

The method described in the following sections could be applied to other archives and extended back in time, as we will discuss in Section 5.

The Nexis UK archive of regional newspapers contains records of the print versions of 568 newspapers from across the United Kingdom (England, Wales, Scotland and Northern Ireland). For our purposes, we focus on the information that can be extracted from them to enrich the NLD which covers just Great Britain (England, Wales, and Scotland).

Whilst Nexis UK coverage is continuous from 1998 to present (LexisNexis Academic, 2014), some selected newspapers have records going back further, although Deacon (2007) cautions that there are some small inconsistencies in how data have been archived. For storage reasons, the Nexis UK archive does not include any original photographs from the news story, so some potentially useful information is lost (Weaver and Bimber, 2008).

Although national newspapers are also archived within Nexis UK, we decided to focus efforts on UK regional newspapers rather than national ones. By their nature, most landslides are local events with local impacts that would be newsworthy at a local to regional level. Any landslides large enough (or with extensive enough impacts) to make the national news would most likely also be captured in the regional press.

At the time of undertaking this research, the BGS had already used media sources (e.g., through Meltwater) to add information to the NLD. However, there are distinct differences between the media sources used by the BGS and the large archive of regional newspapers, Nexis UK, proposed here. Although both sources are digital online services, Meltwater is a record of online news, whereas Nexis UK is a record of printed news. Even if both Meltwater and Nexis UK return records from the same newspaper, the content and length of the stories may vary. In an example given by Greer and Mensing (2006), a study comparing coverage of a news story about genetic cloning across three national broadcast news websites and three national newspaper websites, researchers found that online news stories were generally 20–70% shorter, around 50% of stories were written by newswire services (compared to 10% in print) and generally the websites contained fewer citations. It is not clear how many of the regional newspapers included in Nexis UK database also have an online outlet that is being searched by Meltwater, but it is clear that the content may well differ between the two, and as we will show, the Nexis UK database adds a large number of 'new' records of landslides to the NLD.

3. Methodology

In this section, we present our methodology for searching the Nexis UK archive of regional newspapers to enhance the NLD. This process involves five major steps (Fig. 1):

A. *Construct a set of Boolean search terms* to query the Nexis UK archive (outlined in sub-steps A1–A5).
B. *Apply the search terms* to obtain all articles from a given time period to return a corpus of potentially relevant articles.
C. *Skim-read each article* from this corpus to identify those which are relevant.
D. *Identify whether relevant article refers to a landslide already recorded within the NLD.*
E. *Extract and code relevant information from the relevant articles.*
F. *Pass information on to BGS for quality assurance, cross-checking and NLD data upload.*

3.1. Step A: construct search terms

Nexis UK returns newspaper articles based on a Boolean keyword search (i.e., whether a word or combination of words does or does not appear within an article). There were multiple criteria for the search:

(i) *Maximise the number of articles about landslides in Great Britain*, particularly those that are lesser-known or unlikely to be recorded in the NLD.

(ii) *Minimise the number of false positives* (e.g., articles where the search terms appear in other irrelevant contexts such as "a landslide victory").

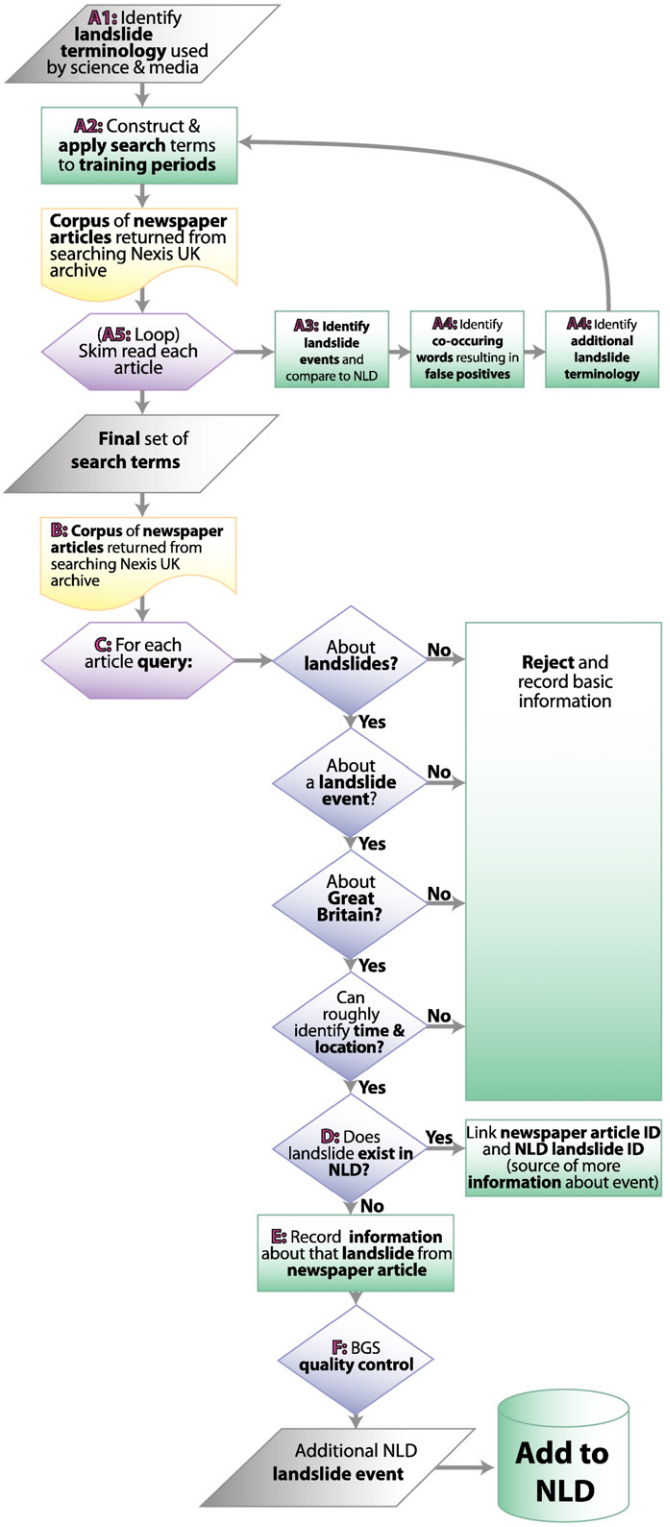

Fig. 1. Flowchart showing the principal stages of the methodology to create a systematic search of Nexis UK regional newspaper articles and extract information about landslide events to add to the richness of the British Geological Survey (BGS) National Landslide Database (NLD). The steps denoted in letters and numbers (A1 to A5, B to F) correspond to the steps identified in the text.

(iii) *Ensure search terms capture any regional or temporal variation in landslide terminology*. For instance the Oxford English Dictionary (OED Online, 2014) notes that landslip is used chiefly in British English and thus would be a less appropriate search term for other parts of the world. Words also change over time, for example the word "slough". The OED notes one meaning for slough (pronounced slaʊ) comes from old English and connotes soft, muddy ground or mires, and another comes from Middle English (pronounced slʌf, sometimes also spelled sluff) and meant outer skin or peel. It was extended metaphorically by 19th century geologists to describe the surficial material shed by engineered embankments and steep scree slopes. In verb form the two meanings come together, insofar as the sloughing of rock or soil is usually *down* into a hole or depression.

Landslides are referenced using many different words by scientists, practitioners and the public, thus we use several Boolean search terms. To refine search terms satisfying the criteria listed above, five sub-steps were completed within Step A:

Stage A1 *Identify key landslide terminology from the sciences and the media.*
Stage A2 *Apply search based on A1 for selected training periods.*
Stage A3 *Read through all resulting articles* from Step A2. Identify landslide events and compare these to those already existing within the NLD.
Stage A4 *Identify any additional terms* used to refer to landslides as well as co-occurring words associated with false positives.
Stage A5 *Incrementally add the additional search terms* found in Step A4 to the existing search terms in A2 and re-apply search. At each stage verify if any articles about landslide events are being filtered out and/or a large number of false positives are being added in.

3.1.1. Search term construction Stage A1 [identify key landslide terminology]

Key landslide terminology outlined by Varnes (1978) and Cruden and Varnes (1996) were assessed, selecting the terms that are more commonly used in the English language or styles of landslide particularly prevalent in Great Britain (Table 1, highlighted in **bold underline**). More commonly used terms known to be used in the British media were added, based on previous BGS experience of searching media, including "landslip", "slope failure" and "slope instability".

3.1.2. Search term construction Stage A2 [apply search based on A1 for selected training periods]

To test the robustness of the combination of search terms from Stage A1, they were applied to the Nexis UK archive of newspaper articles over four sample training periods: 1–31 December for 2004, 2005, 2006 and 2012. Landslide events during December 2012 had a high media profile, with events routinely recorded from national press and social media. During the years 2004 to 2006, 'live' data collection and recording of events were not so systematic, in addition to engineered slope failures and smaller events being rejected. These particular time periods were therefore chosen in order to test semantic variability over the range of the Nexis UK archive and to compare with and add richness to the NLD.

3.1.3. Search term construction Stage A3 [read through all resulting articles from Step A2 to identify landslide events]

Each newspaper article was skim-read to check whether it satisfied the following criteria:

- Is the article relevant (i.e., related to the geomorphic process of landslides)?
- Is the article about a landslide 'event' (rather than general discussion of landslides)?
- Is the article about a landslide event that occurred in Great Britain?
- Is it possible to roughly locate and date the landslide event (possibly requiring further desk-based research)?

If any of the four criteria above were not satisfied, the article was rejected and basic information about the article systematically recorded (to allow future database interrogation). If all of the four criteria above were satisfied, a search of all landslides already existing in the NLD was performed to check whether the landslide was already recorded. If the landslide was already recorded, the newspaper article ID was linked to the NLD landslide ID as a potential source of more information and confirmation. If the landslide event was not in the NLD, as much information as possible about the landslide was extracted from the article and systematically recorded using the same structure as the existing NLD (described in Section 2.2).

3.1.4. Search term construction Stage A4 [identify any additional search terms]

All articles referring to landslides were read carefully to identify any additional terms for landslides used within the texts. This resulted in the additions "cliff collapse" and "land movement". Variations of "cliff collapse" were also added in ("coastline collapse" and "cliff fall"). We also identified co-occurring words associated with false positives (i.e., articles about electoral rather than geologic landslides); all irrelevant articles were coded into themes, and key words selected based on these themes to modify the Boolean filter to remove any articles containing the words "elect" (or derivatives such as elected), "victory", "win", "won", "majority", "submarine" and "porn".

3.1.5. Search term construction Stage A5 [incrementally add the additional search terms]

At each stage, the search of Nexis UK for the training periods was reapplied, and the resulting articles checked to verify that (i) no landslides previously identified were now being filtered out and (ii) no large number of false positives were being added in.

In this Step A5, constructing the final set of search terms used in the rest of our research, there were two cases where a large number of irrelevant articles were returned. The decision was made not to filter results because this would inevitably filter out relevant articles. The first of these was "cliff falls", which captured reports about people falling from the top of cliffs as well as ones about the coastal cliff instability. Given the semantic overlap between these two reporting themes, automated methods could not distinguish between them easily, so it was

Table 1
Landslide terminology for different styles of landslide in various materials from the Varnes (1978) and Cruden and Varnes (1996) classification system. Highlighted in **bold underline** are terms we deemed to be more commonly used in the English language and/or styles of landslide commonly seen in Great Britain.

	Rock	Debris	Soil
Fall	**Rockfall, rock fall**	Debris fall	Soil fall
Topple	**Rock topple**	Debris topple	Soil topple
Slide (rotational)	Slump	Single/multiple/successive	Single/multiple/successive
Slide (translational)	Block slide	Block slide	Slab slide
Planar	Rock slide	**Debris slide**	**Mudslide, mud slide**
Lateral spreading	Rock spreading	Debris spread	Soil spreading
Flow	Rock flow	**Debris flow**	**Mudflow, mud flow**

decided to use manual ones instead. The second case included articles about landslide events occurring abroad (e.g., following a typhoon or earthquake in Asia). Nexis UK offers some additional search filters, such as searching by geography (articles tagged as referring to a specific country) and newspaper section (e.g., only returning articles in the "News" or "Music" sections). However, we chose not to use these filters as sample testing showed that regional newspaper articles are not consistently classified in Nexis UK, therefore the results were too limiting. Manual filtering was used to deal with articles from regional newspapers in Northern Ireland, so as to only choose stories that referred to landslides in Great Britain.

The final search terms that we used for *all* subsequent searches are given below. This includes the use of Boolean logic (OR, AND, NOT) and wild cards (*, !) to search for different derivatives of given terms (e.g., landslide* returns the words landslide, landslided, landslides):

[**landslide*** OR **landslip*** OR **slope failure** OR **rock fall** OR **rockfall** OR **mudflow** OR **mud flow** OR **cliff fall** OR **slope failure** OR **slope instability** OR **debris flow** OR **land movement** OR **cliff collapse** OR **mudslide** OR **mud slide** OR **coastline collapse** OR **rock topple** OR **debris slide** AND NOT (*elect!* OR *victory* OR *win* OR *won* OR *majority* OR *submarine* OR *vote* OR *porn*)]

Terms in **bold underline** indicate that if one instance of that term appears, then the article will be flagged as a potential landslide relevant article. Terms in *italics* indicate that if an article contains any of the bold-underlined black words but also contains one of the italicised words, the article will be filtered out of search results. * = wildcard of 1 character; ! = wildcard of 1 or more characters.

Fig. 2 shows results from applying the final search terms (Step A5, Section 3) to the four training periods (Decembers 2004, 2005, 2006 and 2012). For the December 2004 and 2005 test periods, the NLD did not have any records of landslide events, whereas 4 landslides were identified in each month using the Nexis UK archive. This demonstrates the potential value of applying the method outlined here to enrich the NLD for the period prior to 2006 period when the BGS entered its 'contemporary' phase of data collection. For the December 2006 test period, the NLD contains records of 7 landslide events, 3 of which were also identified in Nexis UK articles. In that month, we also detected 4 additional landslide events not previously recorded in the NLD, representing a 57% increase in database entries for December 2006 by using the Nexis UK archive as an additional source of information. December 2012 was part of a particularly wet season, resulting in many more reported landslides than usual (Pennington and Harrison, 2013). At the time of performing this research, there were 75 landslides in the NLD for December 2012. Of these, 18 events were also identified in the Nexis UK archive. We also detected an additional 6 landslides not recorded in the NLD, increasing the total number of landslide events recorded for December 2012 in the NLD by 8%. The decline between 2006 and 2012 in the proportion of landslides detected using the Nexis method but not currently existing in the NLD, can be explained by the addition of social media as a source of information and the subsequent inclusion of engineered slope failures in the database.

In December 2012, there appear to be proportionally more events (57/81, i.e. 70%) in the NLD that were not found in Nexis UK than in December 2006 (4/11, i.e. 36%). This contrast was investigated for the December 2012 test period by examining the source of information for each landslide event that was found in the NLD but not in the Nexis UK newspaper archive. Fig. 3 shows a breakdown of these sources. The principal reason for these landslide events being in the NLD but not Nexis UK was that they were reported in the media after 31 December 2012. There is good reason to expect that many of these December 2012 events would have been detected using the Nexis UK archive, if instead of searching for a single test month, the time horizon for searching had been extended to overcome this lag time between an event occurring and a story being published about it. The second most frequent

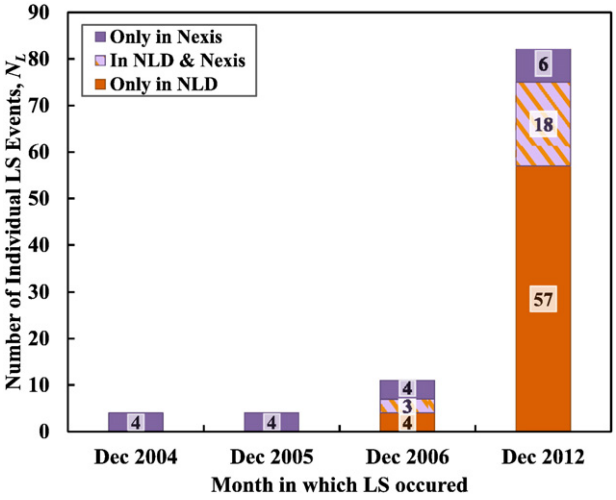

Fig. 2. Number of landslide (LS) events found from the Nexis UK archive search for the December 2004, 2005, 2006 and 2012 training periods that were: (i) only found in Nexis UK (i.e., they did not already exist in the NLD) (solid purple), (ii) found in both the NLD and Nexis UK (hashed purple and orange); and (iii) found in the National Landslide Database (NLD) but not found in the Nexis UK archive (orange).

reason that we found for landslides not being identified in Nexis UK is the source being an online newspaper article from the Newsquest Media Group. This group publishes some 300 local/regional newspapers, but only the print version of many of these newspaper titles is available to search in the Nexis UK archive. From our experience, the content and frequency of publishing vary considerably between the online and print versions. For instance, online news articles may be uploaded daily, whereas the paper is printed once per week, and neither the online nor print version contain all stories of the other, leading to discrepancies in the search results we generated using the Nexis UK method and the media scans provided to the BGS for the NLD by the Meltwater method.

There were a small number of cases where the source was available in the Nexis UK archive, but the specific article was not. This was confirmed by performing additional searches of Nexis using the title of the article and just searching a specific source. This can sometimes happen with freelance or newswire stories where the newspaper does not own copyright and cannot make it available for searching in Nexis UK (LexisNexis Academic, 2014). The majority of the remaining landslide events not identified in the Nexis UK archive search were from sources not available to search in the Nexis archive (e.g., social media, websites). None of the landslide events recorded in the NLD but not returned from the Nexis UK archive appeared to be caused by filtering/errors with the search terms. Although it is not possible to validate these results against the 'true' number of landslides that actually occurred in Great Britain in this period, it does appear that the search terms and method used here has relatively good agreement with existing records in the NLD and is also able to add richness by identifying additional landslide events.

We did not identify any particular regional or temporal variations in landslide terminology. However, all test periods are relatively recent. It is possible that if the search was applied to more historical archives that spatial or temporal trends may appear in the landslide terminology used.

3.2. Step B: apply search terms

The search terms (Step A5, Section 3) were applied to two time periods in the Nexis UK archive: all articles published between 1 January and 31 December for both 2006 and 2012. Once the search was applied, all newspaper articles were downloaded and input into a database to aid categorisation, creating a corpus of potentially relevant stories (see Table 2 for the metadata recorded from two newspaper examples).

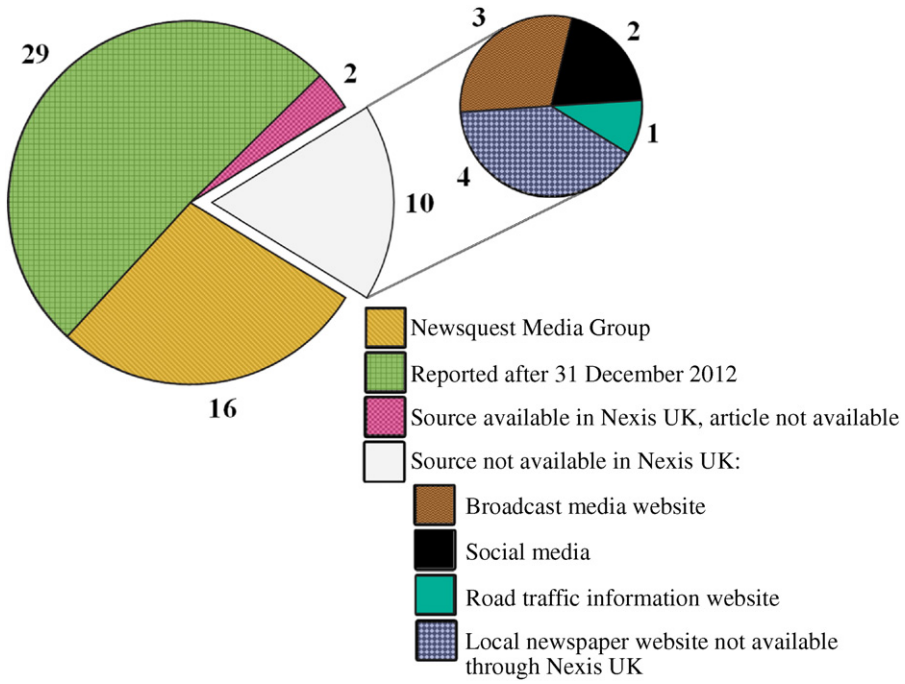

Fig. 3. Breakdown of sources for 57 landslide events noted in the British Geological Survey (BGS) National Landslide Database (NLD) but not in Nexis UK for the December 2012 training period. Also included are the 29 landslide events that occurred in December 2012, but were not reported until after 31 December 2012.

3.3. Step C: skim-read results

The title of each article was skim-read to ascertain whether it was relevant. This is demonstrated in Table 2 where article 1 on Fleetwood Mac is clearly irrelevant from the title and is thus rejected and categorised as "I" (Irrelevant). If the title suggests the article could be relevant, the full text was read to locate and date the possible landslide. In some cases, further desk-based research was required to ascertain whether the article truly referred to a landslide event or not. For example, one newspaper article referred to a landslide but then further described the event as "a building collapsed into a construction site". In such examples, desk-based research was undertaken to identify the exact location of the event using tools including Google Earth time lapse imagery, Google Street View, property websites, social media and other online sources to identify whether this was a landslide, a sinkhole, an issue with slope excavation or another type of event.

3.4. Step D: identify whether relevant article refers to event(s) in the NLD

As detailed in Step A3, if a relevant article contained enough information to approximately locate and date, a search was performed upon the existing NLD to see whether a record of the landslide existed. If so, the article was linked by ID to that landslide event, creating additional confirmation of this event and a potential source of further information to be processed at a later date. Newspaper articles containing more precise information (e.g., improved spatial precision), were used to update the original landslide event.

3.5. Step E: extract and code relevant information from the article

If the landslide did not exist in the NLD, as much information as possible was extracted from the article and categorised according to the BGS NLD pro-forma and a case-by-case judgement of the precision of that information made. An example article is shown in Fig. 4.

4. Results

In this section, we present the results of applying the Nexis UK search method to all regional newspaper articles contained in the database published during the calendar years of both 2006 and 2012. In Section 4.1, we present the overall results of the search before detailed analysis of individual landslide events is undertaken. We then describe how this method adds richness to the NLD through finding previously undetected events (Section 4.2) and the addition of information to existing events (Section 4.3). Finally, in Section 4.4, we discuss the precision to which this information can be estimated from newspaper articles. In Section 5, we will discuss the reliability of this information and potential further applications of the method.

Table 2
Examples of metadata for two newspaper articles returned from searching the Nexis UK archive of regional newspapers for articles published in December 2012. Articles are categorised depending on their relevance and whether they already exist in the NLD or not (N = No, Y = Yes).

Month, year	Article ID	Title	Relevant?	Article type	Enough info to locate?	Landslide in NLD?	Event ID in NLD	New event ID	Full text
Dec-12	1	Don't stop? "Fleetwood Mac will tour until we drop dead" says Stevie Nicks	N	I (irrelevant)	–	–	–	–	"Do not call it a comeback and don't even think of it as a farewell tour. After more than four decades making music and a 2010 tour, Fleetwood Mac will hit the …"
Dec-12	2	Great Christmas getaway hit as signal failure causes widespread disruption	Y	LE (landslide event)	Y	N		N_2012_DEC1	"…the express service made additional calls. In Lancashire, trains between Liverpool and Manchester were diverted because of a landslip near Warrington. Many …"

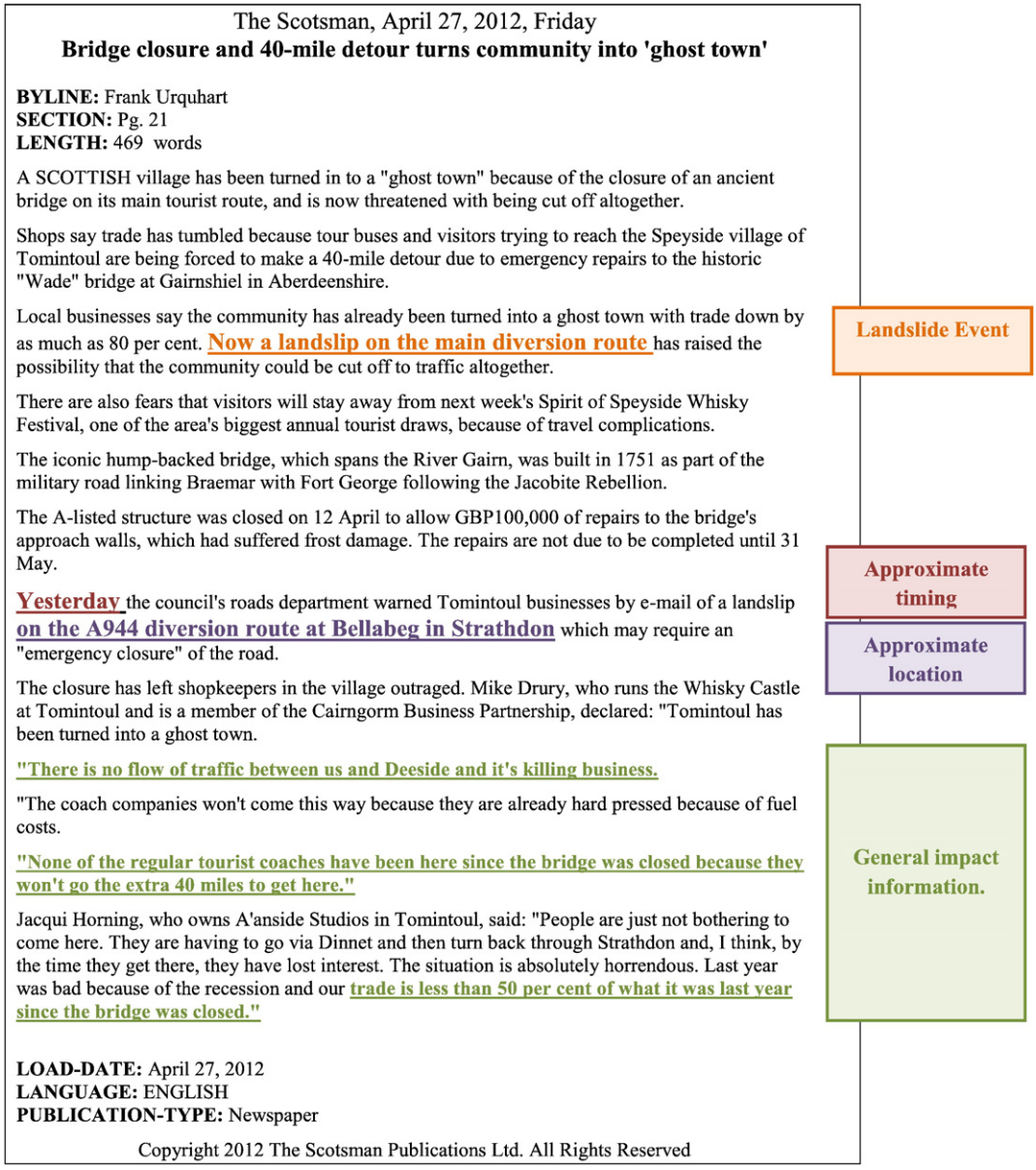

Fig. 4. Example of a newspaper article returned by searching the Nexis UK archive of regional newspapers for articles published in April 2012 and using the search terms given in Step A5, Section 3.1.5. This article refers to a landslide event and gives approximate information about the date, location and impact of that landslide. Newspaper text courtesy of The Scotsman Publications Ltd.

4.1. Search results by type of article

The Nexis UK regional newspaper archive was searched using the terms listed in Step A5, Section 3.1.5 for all articles published between 1 January and 31 December during 2006 and 2012. The initial search (Step B, Section 3.2) resulted in 711 articles in 2006 and 1668 articles in 2012. All articles were then skim-read and categorised into broad types (Step C, Section 3.3), which are listed in Fig. 5. For both periods, around 20% of articles were categorised as completely irrelevant (i.e., false positives), and around 20% of articles were categorised as "general landslide discussion", meaning they referred to landslide phenomena but were not specifically about any particular landslide event.

Broadly, there was a decline in the number of articles discussing landslide events abroad (outside of Great Britain) and historical landslides (those occurring before 2006) between 2006 and 2012. This is countered by an increase in the proportion of 'relevant' articles referring to a landslide event occurring in Great Britain, which rose from 18% in 2006 to 42% in 2012. This is possibly due to the fact that 2012 was a record year for landslides in Great Britain, resulting in increasing public and media interest (Pennington and Harrison, 2013). There was also an increase in the number of articles discussing landslide related policy in 2012. This is largely attributable to relatively unusual high-impact events occurring in 2012, such as fatalities, region-wide railway delays and repeated closure of stretches of road such as the A83 road at Rest and Be Thankful (Scotland), resulting in questioning from the press about what should be done to prevent landslides from a policy perspective. A similar effect has been noted in post-flood event coverage (Escobar and Demeritt, 2014).

Relevant articles referring to landslide events in Great Britain were then analysed more closely to associate them with particular landslides and extract information about those events with which to enrich the NLD in two ways:

(i) Adding landslide events not previously recorded in the NLD
(ii) Capturing more information about landslide events already in the NLD.

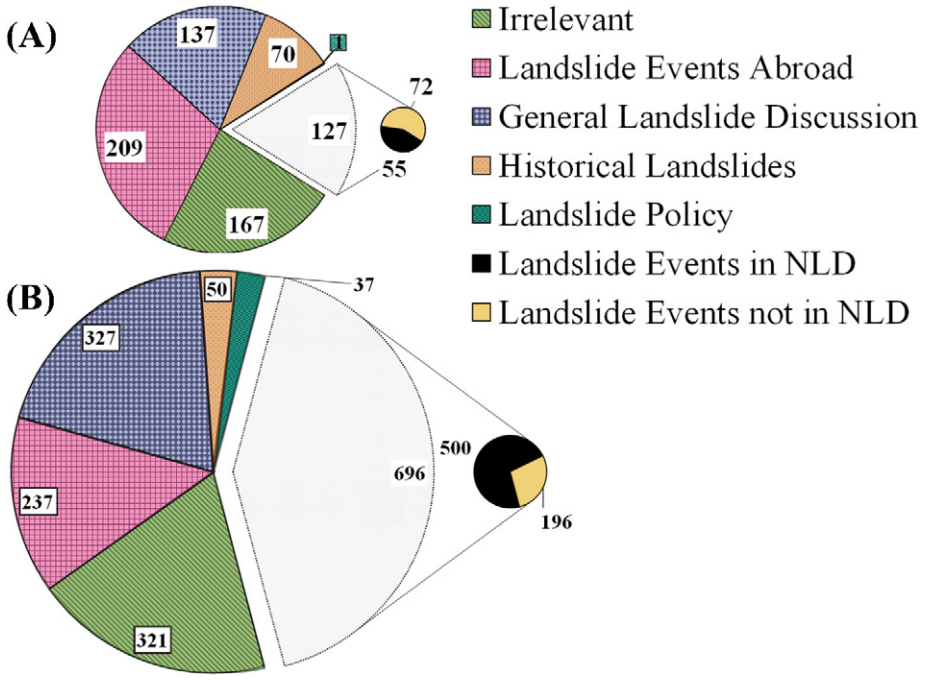

Fig. 5. Number of newspaper articles returned by searching the Nexis UK archive of regional newspapers broken down by type (e.g., articles that are irrelevant, or those that contain relevant information about a landslide event). (A) Results from 711 articles returned from the search of articles published between 1 January 2006–31 December 2006. (B) Results from 1668 returned from the search of articles published between 1 January 2012–31 December 2012.

In the following Section 4.2, we discuss these two ways of enriching the NLD, starting with (i) additional events and their spatial patterning before turning to (ii) the additional information that our method of searching Nexis UK can generate about events already recorded in the NLD.

4.2. Adding landslides to the NLD

Although using Nexis UK we found 268 news articles referring to landslides not previously recorded in the NLD (for calendar years 2006 and 2012), many of these articles were referring to the same, rather smaller, subset of events. Once this repetition in our corpus of articles was accounted for, the final number of additions to the NLD was 39 events for 2006 (compared to 32 events already in the NLD) and 72 events for 2012 (compared to 178 events already in the NLD). This represents a 122% and 40% increase in the number of landslide events in the NLD for 2006 and 2012, respectively. We attribute these NLD additions principally to more and different sources now being searched, along with the majority of new landslides being relatively small in size and thus only of interest to the community in the immediate vicinity. Fig. 6 shows the number of additional landslide events per month for both years. In both years, the seasonal temporal trend in number of landslides per month is roughly the same: high landslide occurrence in the winter, and also a peak in the mid-summer. The pattern in number of additions from the Nexis UK method appears to vary between the years. In 2006, the percentage increase in number of landslides added to the NLD per month varies between 0% and 600% and there does not appear to be a strong relationship between the number of landslides

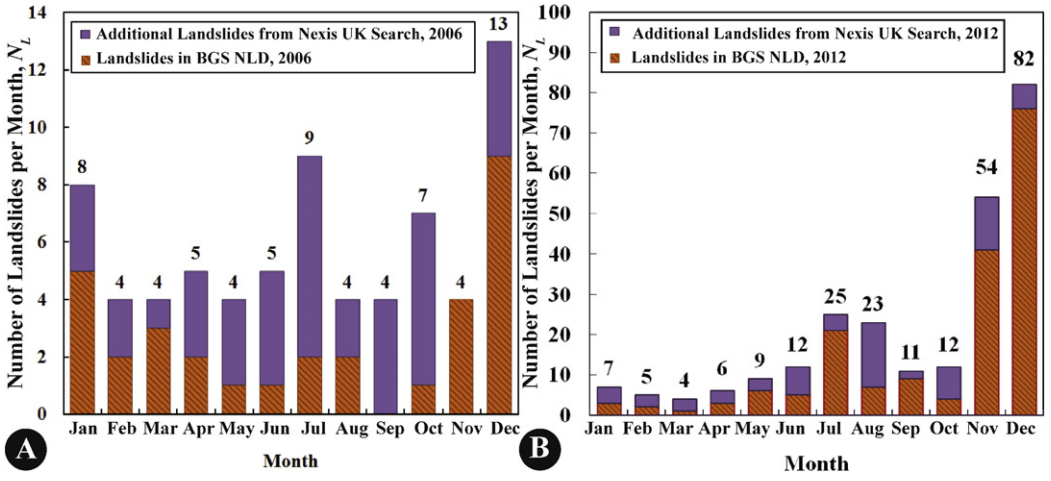

Fig. 6. Number of landslide events per month separated into (i) those already existing in the British Geological Survey (BGS) National Landslide Database (NLD) (orange hashed bars) and (ii) additional landslide events identified by searching the Nexis UK archive of regional newspapers (purple solid bars). Shown are results for 1 January to 31 December for: (A) 2006 (BGS NLD: 32 landslides; additional landslides from Nexis: 39 landslides) and (B) 2012 (BGS NLD: 178 landslides; additional landslides from Nexis: 72 landslides).

already in the NLD and the number of additions. Whereas in 2012, the percentage increase in number of landslides per month varies less (11%–300%) and appears to be weakly linked to the number of landslides already in the NLD for a given month. This suggests that the 'contemporary' phase NLD is a reasonably representative sample of the temporal patterns of landsliding in Great Britain and that the BGS's development of search methods has been effective. Moreover, these results suggest that there is no strong bias for the month landslides are reported in by the media (e.g., in the summer months when there is relatively little political news), although testing of more years of data would be required to confirm this.

Fig. 7 shows separately for 2006 and 2012, the spatial distribution of landslide events already recorded in the NLD at the time of this research, and additional landslides added based on Nexis UK news coverage. The pattern in both years is broadly similar, suggesting no shift over time in the detection biases of this method. The distribution of events previously recorded in the NLD roughly matches that of the additional events detected from the Nexis UK regional newspaper archive but not yet recorded in the NLD. In both 2006 and 2012, both NLD and Nexis UK archive landslides are clustered in the South West of England, with smaller clusters in the North West (Yorkshire Dales), North Wales and the Highlands of Scotland; these areas of significant activity can be directly related to rainfall patterns and topography.

In Fig. 8 we show the spatial distribution of the combined landslides from 2006 and 2012, again for both landslides in the NLD at the time of this research, and additional landslides from Nexis UK, overlaid on a map of landslide susceptibility created from records within the NLD (BGS, 2014c). Broadly, the spatial extent of additional landslides correlates with regions of medium to high susceptibility in the existing susceptibility map.

4.3. Capturing more information about landslides

As well as adding new landslide events to the NLD, the corpus of relevant stories generated by searching Nexis UK was also mined to enrich the NLD by capturing additional information about landslide events. As noted in Section 2.1, the existing BGS pro-forma records >35 attributes (Foster et al., 2008).

For ten landslide events (five in 2006 and five in 2012), additions and amendments were made to the records already in the NLD based on information included in Nexis UK articles. This included more precise dates and locations and additional impact information. Moreover, there are now 55 and 500 additional newspaper articles from Nexis UK for 2006 and 2012 respectively that are linked to individual NLD landslide event entries by ID, acting as additional confirmation for that event and a potential source of further information to be mined at a later date.

Fig. 9 shows a breakdown of the type and/or availability of information available from newspaper articles for each additional landslide event identified from the Nexis UK search ($n = 111$), compared to the types of information available from a subset of the NLD (from 2006–2013, $n = 471$ at the time of doing this research). Newspaper articles are a good source of information for landslide date, approximate location and description of impacts. However, newspaper articles rarely contain more 'geotechnical' information such as the type of landslide, trigger and size. Elliott and Kirschbaum (2007) highlight the difficulty in classifying the type of landslide. Generally, landslide type classification was only possible from the articles in the Nexis UK archive for rock falls, which can be attributed to the relative simplicity of descriptions of large boulders rolling/detaching versus the more visually subtle difference between a planar/rotational slide.

Fig. 9B shows that a trigger for a given landslide event could be identified from newspapers in less than half of cases. Typically the only trigger that could be inferred from an article was heavy or prolonged antecedent rainfall, which articles often described. Our findings based on newspaper articles are broadly consistent with the NLD, which indicates that 63% of landslide events in the NLD in Great Britain were triggered by rainfall. It seems likely that many of the landslides from the Nexis search method missing this triggering information were quite possibly triggered by rainfall.

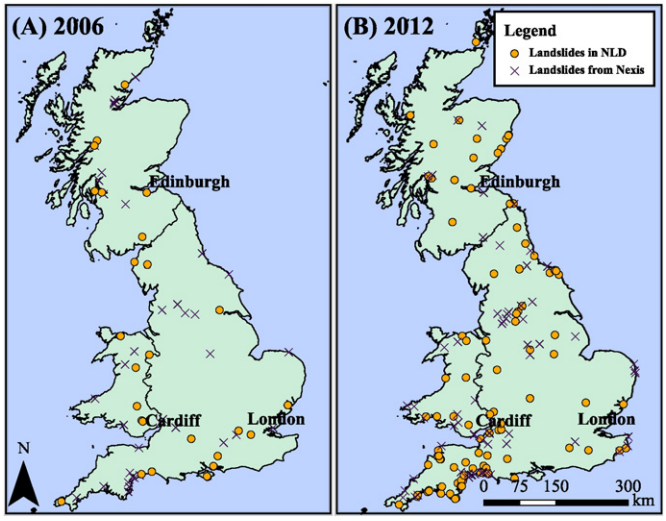

Fig. 7. Spatial distribution of landslides occurring 1 January to 31 December in (A) 2006 and (B) 2012. For both (A) and (B), landslides are separated into (i) all landslides that occurred in 2006 and 2012 that already existed in the NLD at the time of this research (open circles) ($n = 32$ and $n = 178$ for 2006 and 2012 respectively) and (ii) additional landslides identified by searching the Nexis UK archive of regional newspapers (cross symbols) ($n = 39$ and $n = 72$ for 2006 and 2012 respectively).

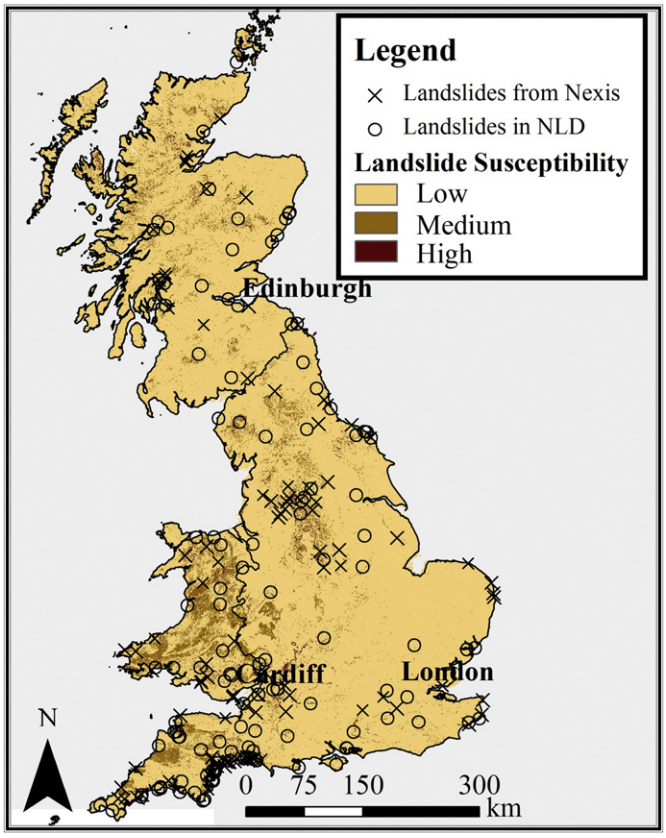

Fig. 8. The spatial distribution of landslides that occurred 1 January to 31 December in both 2006 and 2012 overlaid with a map of landslide susceptibility produced at 1:50,000 scale (BGS, 2014c). Landslides are separated into (i) all landslides that occurred in 2006 and 2012 that already existed in the NLD at the time of this research (open circles) ($n = 210$) and (ii) additional landslides identified by searching the Nexis UK archive (crosses) ($n = 111$).

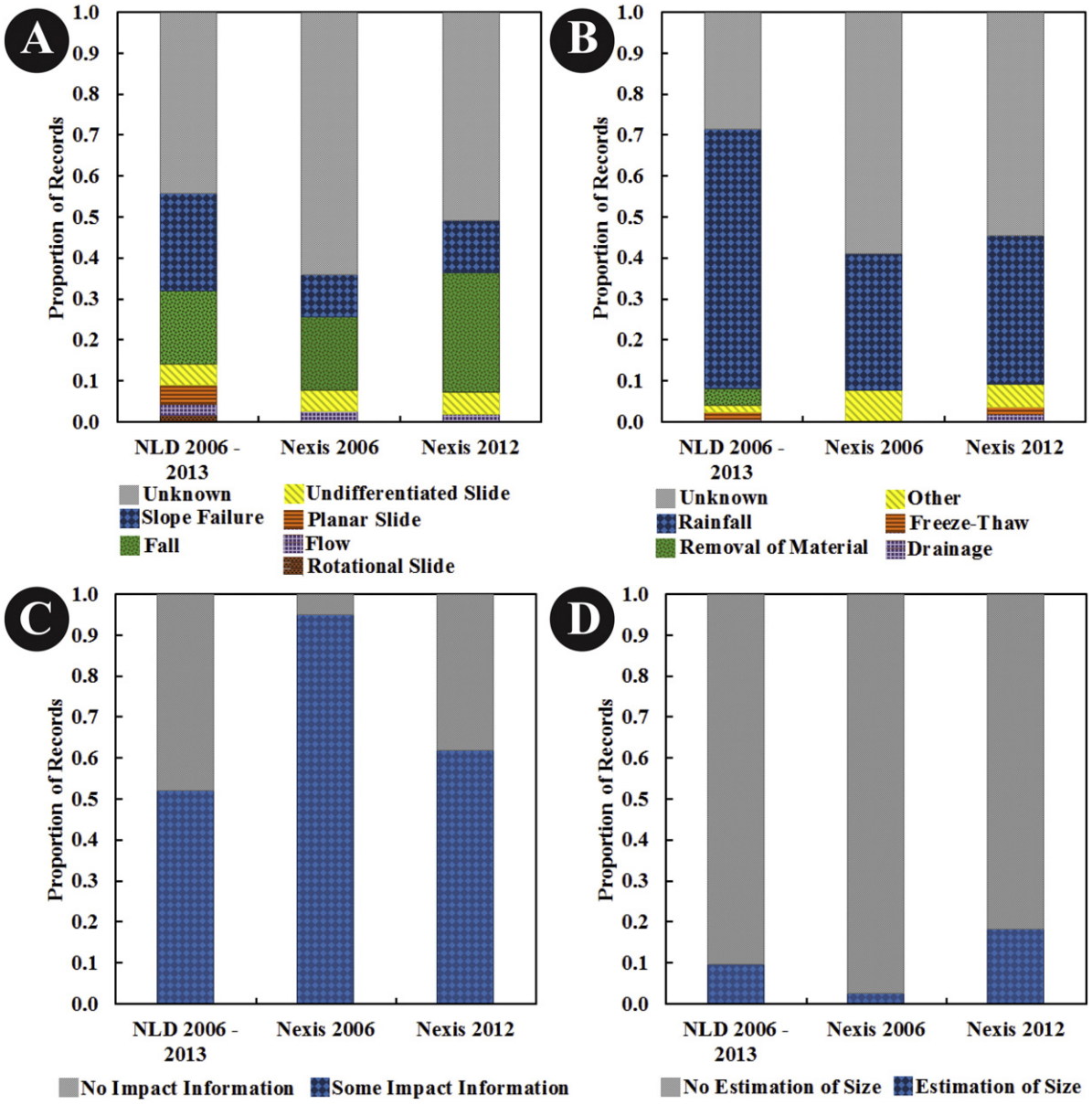

Fig. 9. Type and/or availability of information for each additional landslide event from the Nexis UK archive for 1 January to 31 December of 2006 and 2012 ($n = 111$), compared to the type and/or availability of information available for subset of landslide events already existing in the British Geological Survey (BGS) National Landslide Database (NLD) for the eight year period, 2006 to 2013 ($n = 471$ at the time of this research). (A) Type of landslide using fixed categories based on the Varnes (1978) and Cruden and Varnes (1996) classifications. (B) Trigger of landslide (fixed categories defined by the BGS). (C) Whether any information is available about the impact of the landslide (e.g., fatalities, injured, or other free text). (D) Whether any indication of the landslide size is given (predominantly free text). Panels C and D are presented as binary categories due to the predominantly free-text nature of these data fields.

Newspapers could also be mined for information about the impacts and size of landslides. As these are primarily 'free text' rather than categorical fields in the NLD, results are presented in binary terms of whether information was present or not. Fig. 9C highlights the relative success of extracting landslide impact information from newspaper articles. As mentioned previously, this is most likely due to preferential coverage of landslides that have caused human impact over those that have not. Fig. 9D illustrates that landslide records both from the NLD and newspapers rarely contain information about the size of landslides. Where this information was available, it was generally quoted as a weight in tonnes. Some articles would state the size of a landslide qualitatively (e.g., “small” or “large”), but we did not use these classifications on the grounds that landslide size varies by many orders of magnitude (Stark and Hovius, 2001; Malamud et al., 2004), and truly larger landslides are very rarely seen in Great Britain. Thus, a 'large' landslide to a British journalist may represent a relatively small landslide based on globally observed frequency–size statistics, and even in other British regions might be considered 'medium' or 'small'.

4.4. Assessing the precision of information found using Nexis UK

The precision to which each landslide event can be dated and located from newspaper information was estimated for all additional landslides identified from the Nexis UK archive. Spatial precision (S) is expressed in metres as a radius from the point location (of a given landslide event) given in the database. Date precision (D) is expressed as the amount of time either side of the date given in the database in which the landslide could have occurred. This is generally recorded in categories with increasing units of time (day, week, month, quarter, year). Fig. 10 shows frequency–size plots for the spatial and temporal precision respectively. Approximately 30% of landslide events already existing in the NLD include an estimate of the spatial precision (Fig. 10A). Results are reasonably similar for the 2006 and 2012 periods.

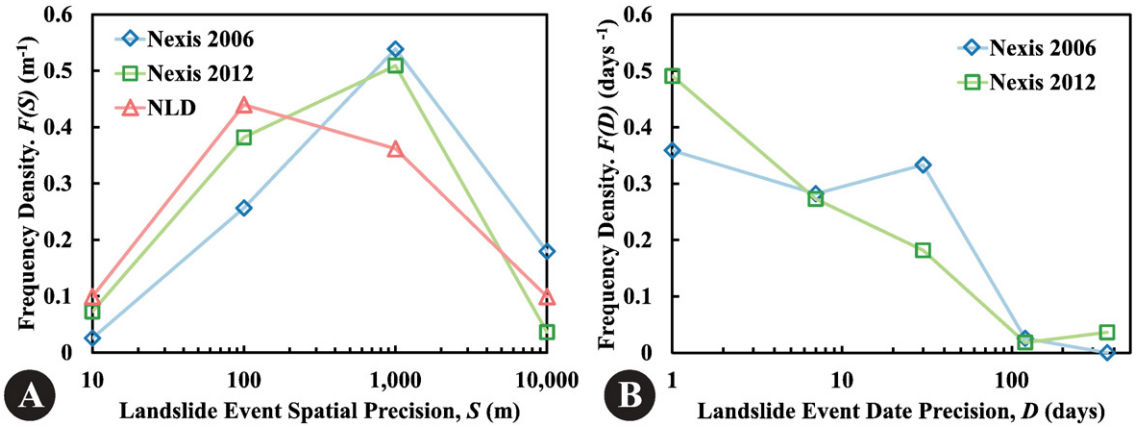

Fig. 10. Frequency density plots of precision of landslide information available from Nexis UK newspaper articles, compared to landslides that already exist in the British Geological Survey (BGS) National Landslide Database (NLD) that occurred between 2006 and 2012 (where data exist). (A) Frequency density of spatial precision of landslide (*x*, *y*) location (*S*, defined as a radius surrounding that point in which the landslide is estimated to have occurred). Estimates of spatial precision are also available for approximately 30% of entries in the NLD, shown with triangle markers. (B) Frequency density of temporal precision of calendar date estimated to be when that landslide occurred (D) measured in days. This estimate of *D* is not included in existing NLD entries.

In both cases the spatial precision of landslide events from the Nexis UK archive is slightly poorer than those landslides already existing in the NLD; in the NLD, the spatial precision peaks at a 100 m radius from the point location of a landslide event, whereas for the Nexis UK, the spatial precision peaks at a 1000 m radius. The date precision (*D*) of additional landslides identified from the Nexis UK archive is generally good (Fig. 10B), with 45% of landslides dated to within 1 day of occurrence and 65–75% of landslides dated to within 1 day to 1 week of occurrence. We hypothesise that this is attributable to a generally short lag between event occurrence and reporting (whilst the event is of public interest).

In Fig. 11, boxplots were used to show the time lag in weeks between a landslide event occurring (estimated from articles) and being reported in Nexis UK newspaper articles, classified by the dating precision of that landslide (see figure caption for details). For landslides where dating precision was within 1 day, the median time lag between the event and reporting is 2 days. For landslides dated within 1 week, month and quarter, the median lag is equal to 1 unit of that time period. For landslides identified in both newspapers and the NLD, an estimate of the date precision is not available, but the median time lag for all these events was 2 days.

5. Discussion

In this paper, we have demonstrated that searching digital newspaper archives is an effective and robust method for adding richness to the NLD. In particular, the search methods we developed were consistently successful in:

(i) Adding previously unrecorded landslide events to the NLD for all but 1 month of the 24 months analysed (Fig. 6, Section 4.2).
(ii) Adding further confidence to many of the existing landslide entries in the NLD by adding additional sources of information (Fig. 2, Section 3.4).

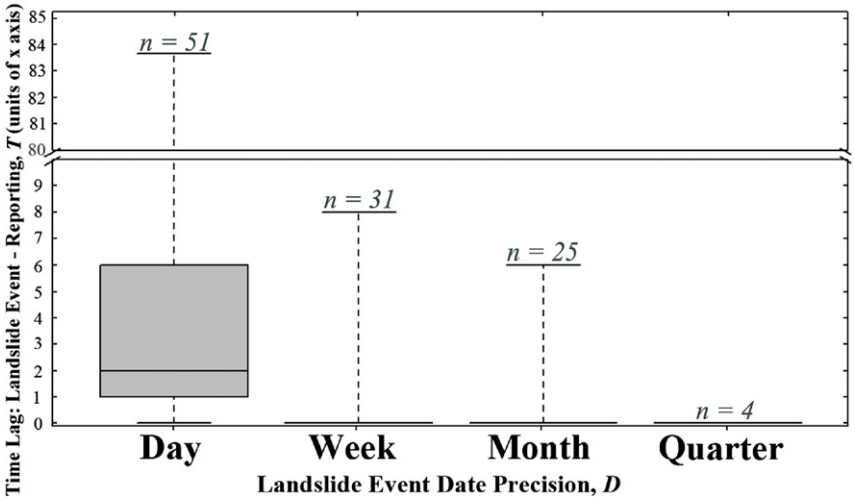

Fig. 11. Boxplot of time lag (*T*) between when landslide is estimated to have occurred and when it was reported (based on article publication date) for the 111 landslide events identified from the Nexis UK archive in 2006 and 2012. Lag is separated by the estimate of temporal precision (*D*) for each landslide and expressed in units of *x*. For example, if a landslide can be dated to within one day of occurrence, the time lag is measured in number of days between occurrence and publication. Similarly, if a landslide can be dated to within a week of occurrence, the time lag is expressed in number of weeks between occurrence and publication. Boxplot whiskers represent the full range of the data for each category.

(iii) Augmenting the information recorded for landslides in the NLD, particularly about their impact (Fig. 9, Section 4.4).

With this proof of concept test, it should now be possible to apply our method to enrich NLD records of historic landslides occurring throughout the period covered by Nexis UK. Moving forward, our search terms could also be applied to supplement the existing sources of information used to alert of BGS of landslide events. This would provide the BGS with a relatively rapid method of 'reconnaissance' to guide whether further investigation (e.g., contact with council/land owner, site visit, remote sensing) may be required.

The most successful element of this work was the addition of landslide events to the NLD. This has resulted in a 122% increase (for 2006) and 40% increase (for 2012) in the total number of landslide events recorded in the NLD. The spatial and temporal distribution, types and triggers for these additional landslides recorded using this method are consistent with existing understandings of landslide susceptibility in Great Britain. These additional landslides also agree broadly with those already recorded in the NLD, which by definition is a 'patchwork' of methods and efforts devoted to data collection strategies (Foster et al., 2012). This agreement provides a basis for added confidence in the NLD as a representative sample of contemporary landsliding in Great Britain, which looks to be growing more complete over time. No single resource will ever provide a complete record of recent landslide events, as events in rural or coastal areas with no impacts are likely to stay unreported, but this research reassures and enhances the current spatio-temporal record. The increasing proportion of events recorded in the NLD relative to those identified from the Nexis UK search highlights the influence of evolving data search-and-capture methodologies. Access to more social media resources, systematic processing and the adaptions of rules regarding the addition of smaller and engineered slope failures has greatly enhanced the 'live' recording of events (Pennington et al., 2015-in this issue).

Beyond this immediate application to enriching the NLD, our paper has wider aims. By outlining in detail a clear methodology for developing and applying Boolean operators for searching digital archives of text data, we have provided earth scientists with a guide for exploiting the new sources of data about earth system processes opened up by the 'digital humanities' and projects like the British Newspaper Archive, which is scanning the vast holdings of historic newspapers held by the British Library to make them available for online searching (British Library, 2014). Following the systematic approach we have described in the paper, it should be possible to develop terms for searching these and other digital archives in order to (i) enrich the records of historic landslides held in the NLD and other landslide inventories (ii) develop similar databases for other hazards.

As with any method, there are uncertainties and biases involved in using such an approach, which we discuss in Section 5.1 along with ways of overcoming them. We then go on (Section 5.2) to discuss how the bias towards events impacting humans could actually be useful in providing a rich source of data for quantifying the costs and other societal impacts of landsliding. In Section 5.3 we go into more detail on how others might extend this research by applying to longer time periods and in its own rights adopting a more automated approach.

5.1. Uncertainties and biases related to the method

Whilst searching newspaper archives offers an effective, relatively low cost method for gathering additional data about landslides and other natural hazard events, there are inevitably uncertainties and limitations to be considered. First, it requires subjective expert judgement to translate journalistic text into the data fields of the NLD. Sometimes relevant information is not explicitly within the news article, but can be inferred, and such inferences can vary between operators (Devoli et al., 2007). In our case, we explicitly used two different people to search the Nexis UK regional newspapers and a one-day training period was performed to ensure consistent interpretation of results. Such 'investigator triangulation' is a well-established method for ensuring the robustness of qualitative research in social science (Baxter and Eyles, 1997).

Second, there are also systematic biases in media coverage that affect its use as a source of landslide inventory data. Media coverage tends to focus attention on large or 'novel' events and those with human interest (Moeller, 2006; Allan et al., 2013) such as an impact on society (e.g., in the UK, road diversions, rail delays, homes being demolished or the closure of coastal footpaths). Also, whilst landslide events are relatively unusual and therefore generally newsworthy, media attention depends on perceptions of salience and if a small landslide occurs on the same day as a large election, the landslide may go unreported, whereas in a period of major landslide impacts (as observed in Great Britain in 2012), landslides may rank high in public interest and receive proportionally more coverage due to an availability bias (Pennington and Harrison, 2013). Thus, although the search strategy used here is systematic, the database we are searching is not a spatially or temporally homogeneous record of events.

5.2. Obtaining information about landslide impact from newspapers

By their very nature, newspaper articles primarily report on "landslides with consequences" (Guzzetti et al., 2003). In a major review of news coverage of disaster events, Quarantelli (1996) found that individual newspapers tend to report on average 90 stories about a particular disaster event, and are most active in the post-event period, providing analytical coverage, resulting in a rich source of information about impacts. In Fig. 9C, we showed that just over 50% of landslide events in the NLD from 2006 onwards contain some information about impact, whereas 60–90% of landslide events identified from the Nexis UK archive contained impact information. Moreover, we found examples of longitudinal reporting of impacts, such as one newspaper article at the time of the event and another article a few months later reporting the remediation works undertaken.

One challenge in compiling records of landslide impacts is defining categories by which impact can be measured. For example, Guzzetti (2000) uses a measure of the number of annual fatalities caused by landslides, Klose et al. (2014) put forward a methodology for measuring the impacts of landslides in economic terms, and Guzzetti et al. (2003) quantify the impact at a regional scale on population, transportation and properties. Schuster and Highland (2003) also note that very few studies consider the impact of landslides upon natural, non-human environments. Because of these difficulties and discrepancies in recording past events, there are few examples in the literature of robust, large-scale forecasting of the impacts of landslides.

Due to the original design and intended research purposes of the NLD, the existing categories in the NLD for recording the impacts of landslides were found to be somewhat insufficient for capturing the rich variety of information available in newspaper articles (see Section 2.2 for a description of categories). Whilst there are fields for number of fatalities, number of persons injured and cost, other impact information is largely recorded as free text. After analysis of Nexis UK articles from 2012 was complete and additional events and information added to the NLD, the list of impact information (for both landslides already existing in the NLD and additions from Nexis UK work) was organised into broad categories, which provide a first indication of the main types of impact observed in Great Britain in a particularly severe year. Fig. 12 shows an infographic of the principal types of impact observed — although it has been noted that the majority of landslides that occurred in 2012 were small shallow failures and in the coming years there may be different types of impact caused by larger, deep seated landslides that have a longer lag time between rainfall and triggering. Nonetheless, this impact information from 2012 now provides a baseline for comparison to other hazard

impact data recording structures (see de Groeve et al. (2013) for a recent review).

Although there is clearly potential to further mine newspaper articles for information about landslide impacts, there are biases such as overestimations, selective coverage and errors in interpretation of impact that must be taken into account (Freudenburg et al., 1996; Quarantelli, 1996). Typically, this would be countered by using the statements made from a range of articles. Such 'source triangulation' is well accepted in the social sciences for dealing with these problems (Baxter and Eyles, 1997). However, due to their local nature, we found that 65% of landslide events were reported in only one article and where the event appeared in multiple articles, the information contained was often repeated verbatim. Nevertheless, newspaper reports can act as a near-real time alert that an impact has occurred and may need to be further investigated (Petrucci et al., 2010).

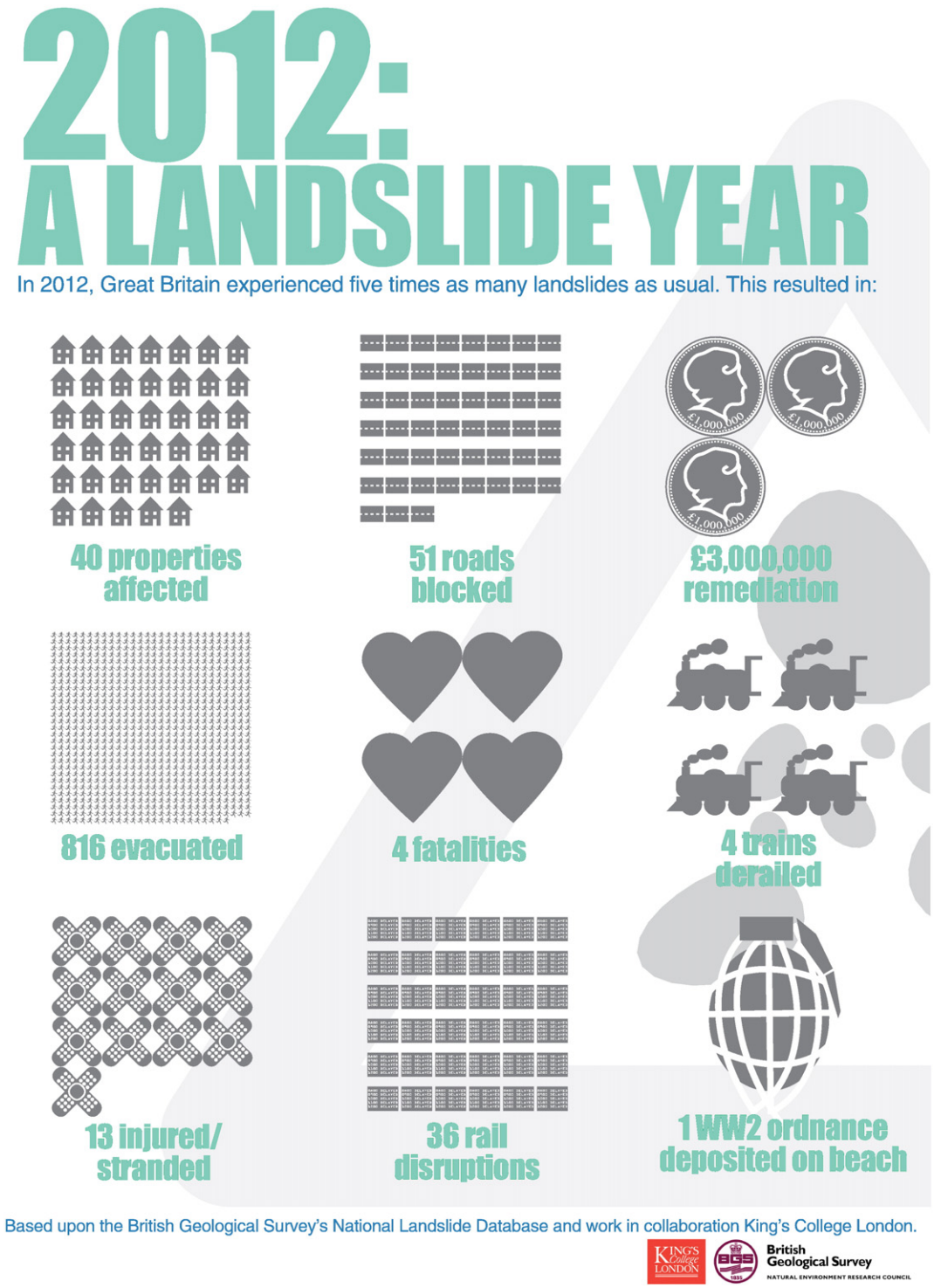

Fig. 12. Infographic of the main types of impacts caused by landslides in Great Britain in 2012. Data from landslide events in the British Geological Survey (BGS) National Landslide Database (NLD) and additional events added in from searching the Nexis UK regional newspaper archive. The final category (WW2 ordnance deposited on beach) represents other more irregular or chaotic impacts.

5.3. Potential extensions to the method

As already described above, we are not the first to use newspaper articles as a source of information about landslide events. Newspaper have also been successfully drawn on as a major source of information about historical events (e.g., Guzzetti et al., 1994; Elliott and Kirschbaum, 2007; Petley, 2012) and to supplement other landslide inventories (e.g., Miller et al., 2009; Pradhan and Lee, 2010). Although these studies have undoubtedly been performed with attention to detail and in a systematic way, there is relatively little discussion within the literature of the detailed process of constructing a robust search strategy with the aim of capturing as many relevant articles as possible. It is hoped that by detailing the methodological steps involved and addressing related issues of uncertainty, this paper will make it easier for others to apply this method. We now discuss three potential extensions to the method we have explored in this paper: (i) extend archival searching farther back in time, (ii) increase speed and automation of the archival searching, and (iii) extend archival searching method for landslides to other countries or other hazard databases.

(i) Extend archival searching farther back in time.

To produce high quality landslide susceptibility maps and broadly have a good understanding of the landscape setting in which landslides occur across a region, we often require multi-temporal inventories of landslides, extending back over a number of decades. This is an issue for retrospective studies, as many landslides (particularly smaller ones) are 'erased' from the landscape via erosional processes within a few months to years (Malamud et al., 2004; Bell et al., 2012). Thus, to produce historical inventories, we often rely on records of landslides from proxy sources. Indeed, in perhaps the best example of a long-term (~90 years) archive of landslide events (The Italian AVI Project), over 60% of records of landslide events come from newspapers, and the others from reports and interviews (Guzzetti et al., 1999). Other examples include a database of historical landslides occurring in Utah from 1850 to 1978 (Elliott and Kirschbaum, 2007) and landslides occurring before 1990 in Nicaragua (Devoli et al., 2007). Although the Nexis UK archive only extends back to 1998, there have been many advances in the digitisation, character recognition and compilation of historical UK newspaper sources going back considerably further, suggesting that this method could be applied to much longer time periods to gain a better long-term understanding of landslide phenomena. For example, the British library has been undertaking a project to digitise its archive of newspapers extending back to 1800 (British Library, 2014). It is likely that the search terms listed in Section 3 would need to be adjusted to take into account historical variations in terminology, but this presents an opportunity to gain further insight into landsliding in Great Britain over a relatively long timescale.

(ii) Increase speed and automation of the archival searching.

There have been considerable developments in the field of automated newspaper content analysis using computers to identify the meaning of sentences within a text and extract information into a database; and this has been applied to fields such as political science (van Atteveldt et al., 2008; Hopkins and King, 2010), economics (Sprenger and Welpe, 2011) and the policy dimensions of environmental phenomena such as hurricanes (Soroka et al., 2009) and climate change (Kirilenko and Stepchenkova, 2012). This could be of use to more rapidly process the large number of articles returned and retrospectively populate the database over longer time periods, particularly in countries where a large number of landslides occur annually. There are questions, however, about how easily this automated approach could be adapted to the creation of landslide event databases due to the indirect descriptions of events and the need for additional research to extract information (discussed in Sections 5.1 and 5.2). There have been considerable advances in the ability to automate searches of large volumes of social media, so it is possible that now robust search terms have been developed, it may be possible to apply a more automated approach to the task.

(iii) Extend archival searching method for landslides to other countries or other hazard databases

The issues of database completeness are not specific to the field of landslides in Great Britain. As mentioned in Section 2.1, Van Den Eeckhaut et al. (2012) found that the majority of European countries that maintain landslide databases estimate the completeness to be around 50%. At a global level, Guzzetti et al. (2012) estimated that only around 1% of slopes have associated landslide inventory maps. Yet, detailed, systematic, well-produced landslide inventories are fundamental in both applied risk analysis (e.g., Harp et al., 2011) and scientific research (e.g., Malamud et al., 2004). Indeed, it is acknowledged across many hazard-related disciplines that database incompleteness is an issue, and various proxy records have the potential to fill some of the gaps in our knowledge. Examples include Stucchi et al. (2004) for seismology, Barredo (2007) for flooding and Blackford and Chambers (1991) with respect to climatology.

The method outlined in this paper has demonstrated a good ability to identify small landslides that might otherwise be missed by other methods of inventory production, historical landslides that may have been erased from the landscape and more generally, detailed accounts of hazard impact. The search terms outlined in Step A5, Section 3 could be applied 'as is' to the remaining years of the Nexis UK archive (1998 to present), and perhaps with some further verification of temporal variations in terminology to the British Newspaper archive, which dates back to the 1800s (British Library, 2014). The Nexis archive also contains material from many countries across the globe, and has a similar level of coverage for France, Germany and the Netherlands (LexisNexis Academic, 2014). By clearly outlining the steps involved in search terminology experimentation (Fig. 1), this method can now be applied broadly to other countries or other hazards to create robust, systematic inventories of hazard information from newspaper articles.

6. Conclusions

This paper has set out a method to construct a set of Boolean terms and systematically search the Nexis UK archive of 568 regional newspapers for information about landslide events in Great Britain. When applied to all newspaper articles published in 2006 and 2012, this method added richness to the existing National Landslide Database (NLD) in three ways: (i) Additional landside events were added that had not previously been recorded in the NLD, resulting in a 120% and 40% increase in the number of documented landslides in Great Britain in 2006 and 2012 respectively; (ii) NLD records of landslide events were augmented, by populating more fields of information and also providing additional sources of confirmation to many events, thus increasing the robustness of the database; and (iii) Landslide impact information could be obtained from newspaper reports. There are some issues with uncertainty and inhomogeneities in media coverage of hazard events, which require caution. This method should be considered as supplementary to more robust methods of landslide database production (such as field investigation and remote sensing). Nonetheless, this method represents a relatively quick, low-cost way of identifying events that may require further investigation. In explicitly outlining the steps involved in creating a robust, systematic search, we hope this method can be applied to other landslide and other hazard databases (such as flooding) to increase the richness of past records and thus improve the ability to forecast future events.

Acknowledgements

- Harry Whittle (University of Leeds) for data processing.
- Other members of the BGS Landslide Team for their input and

proofreading.
- BGS for use of the NLD.
- This project was funded by the Engineering and Physical Sciences Research Council (EPSRC) and is part of the Probability Uncertainty and Risk in the Environment (PURE) research programme (Grant Number: IP13-004), funded by the Natural Environment Research Council (NERC) and managed by the Industrial Mathematics KTN.

References

Allan, S., Adam, B., Carter, C., 2013. Introduction: the media politics of environmental risk. In: Allan, S., Adam, B., Carter, C. (Eds.), Environmental Risks and the Media. Routledge, London, pp. 1–26.
Barredo, J., 2007. Major flood disasters in Europe: 1950–2005. Nat. Hazards 42, 125–148.
Baxter, J., Eyles, J., 1997. Evaluating qualitative research in social geography: establishing 'rigour' in interview analysis. Trans. Inst. Brit. Geogr. 22, 505–525.
Bell, R., Petschko, H., Röhrs, M., Dix, A., 2012. Assessment of landslide age, landslide persistence and human impact using airborne laser scanning digital terrain models. Geogr. Ann. A 94, 135–153.
BGS (British Geological Survey), 2014a. BGS National Landslide Database [online]. Available at:, http://www.bgs.ac.uk/landslides/nld.html (Last accessed 02 Sep 2014).
BGS (British Geological Survey), 2014b. BGS "Report a Landslide" web portal [online]. Available at:, http://www.bgs.ac.uk/landslides/report.html (Last accessed 02 Sep 2014).
BGS (British Geological Survey), 2014c. National Landslide Potential Map (GeoSure).
Black, A.R., Law, F.M., 2004. Development and utilization of a national web-based chronology of hydrological events. Hydrol. Sci. J. 49, 237–246.
Blackford, J.J., Chambers, F.M., 1991. Proxy records of climate from blanket mires: evidence for a Dark Age (1400 BP) climatic deterioration in the British Isles. The Holocene 1, 63–67.
Brázdil, R., Pfister, C., Wanner, H., Von Storch, H., Luterbacher, J., 2005. Historical climatology in Europe—the state of the art. Clim. Chang. 70, 363–430.
British Library, 2014. The British Newspaper Archive [online]. Available at:, http://www.britishnewspaperarchive.co.uk/ (Last accessed 02 Sep 2014).
Bromhead, E.N., Ibsen, M.L., 2006. A review of landsliding and coastal erosion damage to historic fortifications in South East England. Landslides 3, 341–347.
Brunsden, D., 1985. Landslide types, mechanisms, recognition, identification. In: Morgan, C.S. (Ed.), Landslides in the South Wales Coalfield: Proceedings, Symposium, Polytechnic of Wales, Pontypridd, pp. 19–28.
Carrara, A., Crosta, G., Frattini, P., 2003. Geomorphological and historical data in assessing landslide hazard. Earth Surf. Proc. Land. 28, 1125–1142.
Carvalho, A., 2007. Ideological cultures and media discourses on scientific knowledge: re-reading news on climate change. Public Underst. Sci. 16, 223–243.
CNR-GNDCI (Consiglio Nazionale delle Ricerche-Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche), 2014. AVI Project home page [Online]. Available at:, http://avi.gndci.cnr.it/ (Last accessed 02 Sep 2014).
Cruden, D., Varnes, D.J., 1996. Landslide types and processes. In: Turner, A.K., Schuster, R.L. (Eds.), Landslides: Investigation and Mitigation. Transportation Research Board, Washington D.C., pp. 36–75.
De Groeve, T., Poljansek, K., Ehrlich, D., 2013. Recording Disaster Losses. Recommendations for a European Approach. Report EUR 26111 EN. European Commission Joint Research Centre, Ispra, Italy.
Deacon, D., 2007. Yesterday's papers and today's technology digital newspaper archives and 'push button' content analysis. Eur. J. Commun. 22, 5–25.
Demeritt, D., 1991. Climate, cropping, and society in Vermont, 1820–1850. Vt. Hist. 50, 133–165.
Devoli, G., Morales, A., Høeg, K., 2007. Historical landslides in Nicaragua—collection and analysis of data. Landslides 4, 5–18.
Domínguez-Cuesta, M.J., Jiménez Sánchez, M., Rodríguez García, A., 1999. Press archives as temporal records of landslides in the North of Spain: relationships between rainfall and instability slope events. Geomorphology 30, 125–132.
Elliott, A., Kirschbaum, M.J., 2007. The Preliminary Landslide History Database of Utah, 1850–1978. Open File Report 514. Utah Geological Survey, Salt Lake City (12 pp.).
Escobar, M.P., Demeritt, D., 2014. Flooding and the framing of risk in British broadsheets, 1985–2010. Public Underst. Sci. 23, 454–471.
Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Eng. Geol. 102, 99–111.
Fischer, H.W., 1994. Response to Disaster: Fact versus Fiction and Its Perpetuation. University Press of America (160 pp.).
Foster, C., Gibson, A., Wildman, G., British Geological Survey, 2008. The new national Landslide Database and Landslide Hazard Assessment of Great Britain. Proceedings of the First World Landslide Forum, Tokyo, Japan, pp. 203–206.
Foster, C., Pennington, C.V.L., Culshaw, M.G., Lawrie, K., 2012. The National Landslide Database of Great Britain: development, evolution and applications. Environ. Earth Sci. 66, 941–953.
Freudenburg, W.R., Coleman, C.-L., Gonzales, J., Helgeland, C., 1996. Media coverage of hazard events: analyzing the assumptions. Risk Anal. 16, 31–42.
Fuchs, S., Heiss, K., Huebl, J., 2007. Towards an empirical vulnerability function for use in debris flow risk assessment. Nat. Hazards Earth Syst. Sci. 7, 495–506.
Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008. Comparing landslide inventory maps. Geomorphology 94, 268–289.
GDACS, 2014. Global disaster alert and coordination system [online]. Available at:, http://www.gdacs.org/ (Last accessed 02 Sep 2014).
Gibson, A.D., Culshaw, M.G., Dashwood, C., Pennington, C.V.L., 2013. Landslide management in the UK—the problem of managing hazards in a 'low-risk' environment. Landslides 10, 599–610.
Glade, T., Crozier, M., 1996. Towards a national landslide information base for New Zealand. N. Z. Geogr. 52, 29–40.
Greer, J.D., Mensing, D., 2006. The evolution of online newspapers: a longitudinal content analysis, 1997–2003. In: Li, X. (Ed.), Internet Newspapers: Making of a Mainstream Medium. Lawrence Erlbaum Associates, London, pp. 13–32.
Guzzetti, F., 2000. Landslide fatalities and the evaluation of landslide risk in Italy. Eng. Geol. 58, 89–107.
Guzzetti, F., Tonelli, G., 2004. Information system on hydrological and geomorphological catastrophes in Italy (SICI): a tool for managing landslide and flood hazards. Nat. Hazards Earth Syst. Sci. 4, 213–232.
Guzzetti, F., Cardinali, M., Reichenbach, P., 1994. The AVI project: a bibliographical and archive inventory of landslides and floods in Italy. Environ. Manag. 18, 623–633.
Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31, 181–216.
Guzzetti, F., Reichenbach, P., Cardinali, M., Ardizzone, F., Galli, M., 2003. The impact of landslides in the Umbria region, central Italy. Nat. Hazards Earth Syst. Sci. 3, 469–486.
Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., Ardizzone, F., 2005. Probabilistic landslide hazard assessment at the basin scale. Geomorphology 72, 272–299.
Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.T., 2012. Landslide inventory maps: new tools for an old problem. Earth-Sci. Rev. 112, 42–66.
Harp, E.L., Keefer, D.K., Sato, H.P., Yagi, H., 2011. Landslide inventories: the essential part of seismic landslide hazard analyses. Eng. Geol. 122, 9–21.
Heim, A., 1989 [1932]. Landslides and Human Lives [Translated by Skermer, N. A.]. BiTech Publishers, Vancouver, B.C., 195 pp.
Hopkins, D.J., King, G., 2010. A method of automated nonparametric content analysis for social science. Am. J. Polit. Sci. 54, 229–247.
Hurford, A.P., Priest, S.J., Parker, D.J., Lumbroso, D.M., 2012. The effectiveness of extreme rainfall alerts in predicting surface water flooding in England and Wales. Int. J. Climatol. 32, 1768–1774.
Ibsen, M.-L., Brunsden, D., 1996. The nature, use and problems of historical archives for the temporal occurrence of landslides, with specific reference to the south coast of Britain, Ventnor, Isle of Wight. Geomorphology 15, 241–258.
Jones, D.K.C., Lee, E.M., 1994. Landsliding in Great Britain. Department of the Environment, London (390 pp.).
Kasperson, R.E., Renn, O., Slovic, P., Brown, H.S., Emel, J., Goble, R., Kasperson, J.X., Ratick, S., 1988. The social amplification of risk—a conceptual-framework. Risk Anal. 8, 177–187.
Kirilenko, A., Stepchenkova, S., 2012. Climate change discourse in mass media: application of computer-assisted content analysis. J. Environ. Stud. Sci. 2, 178–191.
Kirschbaum, D., Adler, R., Hong, Y., Hill, S., Lerner-Lam, A., 2010. A global landslide catalog for hazard applications: method, results, and limitations. Nat. Hazards 52, 561–575.
Klose, M., Damm, B., Terhorst, B., 2014. Landslide cost modeling for transportation infrastructures: a methodological approach. Landslides 12, 1–14.
LexisNexis Academic, 2014. [online] Available at:, http://academic.lexisnexis.com (Last accessed 02 Sep 2014).
Llasat, M.C., Llasat-Botija, M., López, L., 2009. A press database on natural risks and its application in the study of floods in Northeastern Spain. Nat. Hazards Earth Syst. Sci. 9, 2049–2061.
Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslide inventories and their statistical properties. Earth Surf. Proc. Land. 29, 687–711.
Meltwater, 2014. Meltwater news [online]. Available at:, http://www.meltwater.com/products/meltwater-news/ (Last accessed 02 Sep 2014).
Met Office, 2013. Natural hazards partnership: hazard impact model [online]. Available at:, http://www.metoffice.gov.uk/nhp/hazard-impact-model (Last Accessed 02 Sep 2014).
Miller, S., Brewer, T., Harris, N., 2009. Rainfall thresholding and susceptibility assessment of rainfall-induced landslides: application to landslide management in St Thomas, Jamaica. Bull. Eng. Geol. Environ. 68, 539–550.
Moeller, S.D., 2006. "Regarding the pain of others": media, bias and the coverage of international disasters. J. Int. Aff. 59, 173–196.
Moores, A.J., Rees, J.G. (Eds.), 2011. UK Flood and Coastal Erosion Risk Management Research Strategy. Living with Environmental Change (58 pp. Available at: http://www.lwec.org.uk/our-work/uk-first-flood-research-strategy [Last accessed 02 Sep 2014]).
NCSE, 2007. Nineteenth century serials edition [online]. Available at:, http://www.ncse.ac.uk/commentary/index.html (Last accessed 02 Sep 2014).
OED Online, 2014. Oxford English Dictionary. Oxford University Press ([online]. Available at: http://www.oed.com/ [Last accessed 02 Sep 2014]).
Parry, S., Marsh, T., Kendon, M., 2013. 2012: from drought to floods in England and Wales. Weather 68, 268–274.
Pennington, C.V.L., Harrison, A., 2013. 2012 — Landslide Year, Geoscientist. The Geological Society of London, London, pp. 10–15.
Pennington, C., Foster, C., Chambers, J., Jenkins, G., 2009. Landslide research at the British geological survey: capture, storage and interpretation on a national and site-specific scale. Acta Geol. Sin. Engl. Ed. 83, 991–999.
Pennington, C.V.L., Freeborough, K., Dashwood, C., Dijkstra, T., Lawrie, K., 2015. The National Landslide Database of Great Britain: acquisition, communication and the role of social media. Geomorphology 249, 112–119 (in this issue).

Peters, H.P., Brossard, D., De Cheveigné, S., Dunwoody, S., Kallfass, M., Miller, S., Tsuchida, S., 2008. Interactions with the mass media. Science 321, 204–205.
Petley, D., 2012. Global patterns of loss of life from landslides. Geology 40, 927–930.
Petley, D.N., Hearn, G.J., Hart, A., Rosser, N.J., Dunning, S.A., Oven, K., Mitchell, W.A., 2007. Trends in landslide occurrence in Nepal. Nat. Hazards 43, 23–44.
Petrucci, O., Pasqua, A.A., Gullà, G., 2010. Landslide damage assessment using the Support Analysis Framework (SAF): the 2009 landsliding event in Calabria (Italy). Adv. Geosci. 26, 13–17.
Pradhan, B., Lee, S., 2010. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. Landslides 7, 13–30.
Public Health England, 2014. Available online at:, https://www.gov.uk/government/organisations/public-health-england (Last accessed 02 Sep 2014).
Quan Luna, B., Blahut, J., Van Westen, C.J., Sterlacchini, S., van Asch, T.W., Akbas, S.O., 2011. The application of numerical debris flow modelling for the generation of physical vulnerability curves. Nat. Hazards Earth Syst. Sci. 11, 2047–2060.
Quarantelli, E.L., 1996. Local mass media operations in disasters in the USA. Disaster Prev. Manag. 5, 5–10.
Raška, P., Zábranský, V., Dubišar, J., Kadlec, A., Hrbáčová, A., Strnad, T., 2014. Documentary proxies and interdisciplinary research on historic geomorphologic hazards: a discussion of the current state from a central European perspective. Nat. Hazards 70, 705–732.
Salvati, P., Balducci, V., Bianchi, C., Guzzetti, F., Tonelli, G., 2009. A Web GIS for the dissemination of information on historical landslides and floods in Umbria, Italy. GeoInformatica 13, 305–322.
Schuster, R.L., Highland, L.M., 2003. Impact of landslides and innovative landslide-mitigation measures on the natural environment [online]. Available at:, http://pubs.usgs.gov/op/HongKongJuly/HongKongJuly21sm.pdf (Last accessed 02 Sep 2014).
Soeters, R., Van Westen, C.J., 1996. Slope instability recognition, analysis, and zonation. In: Turner, A.K., Schuster, R.L. (Eds.), Landslide: Investigations and Mitigation Special Report 247. Transportation Research Board, Washington D.C., pp. 129–177.
Soroka, S.N., Farnsworth, S.J., Young, L., Lawlor, A., 2009. Environment and energy policy: comparing reports from US and Canadian television news. Annual Meeting of the American Political Science Association, Toronto, Canada, pp. 3–6.
Sprenger, T.O., Welpe, I.M., 2011. News or Noise? The stock market reaction to different types of company-specific news events [online]. Available at:, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1734632 (Last accessed 02 Sep 2014).
Stark, C.P., Hovius, N., 2001. The characterization of landslide size distributions. Geophys. Res. Lett. 28, 1091–1094.
Stucchi, M., Albini, P., Mirto, M., Rebez, A., 2004. Assessing the completeness of Italian historical earthquake data. Ann. Geophys. 47, 659–673.
Tarhule, A., 2005. Damaging rainfall and flooding: the other Sahel hazards. Clim. Chang. 72, 355–377.
Trimble, S.W., 2008. The use of historical data and artifacts in geomorphology. Prog. Phys. Geogr. 32, 3–29.
Tschoegl, L., Below, R., Guha-Sapir, D., 2006. An analytical review of selected data sets on natural disasters and impacts, Centre for Research on the Epidemiology of Disasters. UNDP/CRED Workshop on Improving Compilation of Reliable Data on Disaster Occurrence and Impact, Bangkok, Thailand (21 pp.).
UN (United Nations), 2005. Hyogo Framework for Action 2005–2015: Building the Resilience of Nations and Communities to Disasters, United Nations, Geneva (28 pp.).
van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., 2008. Parsing, semantic networks, and political authority using syntactic analysis to extract semantic relations from Dutch newspaper articles. Polit. Anal. 16, 428–446.
Van Den Eeckhaut, M., Hervás, J., 2012. State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk. Geomorphology 139–140, 545–558.
Van Den Eeckhaut, M., Hervás, J., Jaedicke, C., Malet, J.P., Montanarella, L., Nadim, F., 2012. Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. Landslides 9, 357–369.
van Westen, C.J., van Asch, T.W.J., Soeters, R., 2006. Landslide hazard and risk zonation—why is it still so difficult? Bull. Eng. Geol. Environ. 65, 167–184.
van Westen, C.J., Castellanos, E., Kuriakose, S.L., 2008. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Eng. Geol. 102, 112–131.
Varnes, D.J., 1978. Slope movements types and processes. In: Schuster, R.L., Krizek, R.J. (Eds.), Transportation Research Board Special Report 176: Landslides: Analysis and Control. Transportation Research Board, Washington DC, pp. 11–33.
Weaver, D.A., Bimber, B., 2008. Finding news stories: a comparison of searches using Lexis Nexis and Google News. J. Mass Commun. Q. 85, 515–530.
Wills, C.J., McCrink, T.P., 2002. Comparing landslide inventories: the map depends on the method. Environ. Eng. Geosci. 8, 279–293.
Woolley, R.R., Marsell, R.E., Grover, N.C., 1946. Cloudburst Floods in Utah, 1850–1938. Water-Supply Paper 994. US Government Printing Office (154 pp.).