



# Data expansion: the potential of grey literature for understanding floods

S. Uhlemann<sup>1,2</sup>, R. Bertelmann<sup>3</sup>, and B. Merz<sup>1</sup>

<sup>1</sup>GFZ German Research Centre for Geosciences, Section Hydrology, Potsdam, Germany

<sup>2</sup>University of Potsdam, Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany

<sup>3</sup>GFZ German Research Centre for Geosciences, Library and Information Services, Potsdam, Germany

Correspondence to: S. Uhlemann (uhlemann@gfz-potsdam.de)

Received: 17 September 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 27 September 2012

Revised: 22 January 2013 – Accepted: 12 February 2013 – Published: 4 March 2013

**Abstract.** Sophisticated methods have been developed and become standard in analysing floods as well as for assessing flood risk. However, increasingly critique of the current standards and scientific practice can be found both in the flood hydrology community as well as in the risk community who argue that the considerable amount of information already available on natural disasters has not been adequately deployed and brought to effective use. We describe this phenomenon as a failure to synthesize knowledge that results from barriers and ignorance in awareness, use and management of the entire spectrum of relevant content, that is, data, information and knowledge. In this paper we argue that the scientific community in flood risk research ignores event-specific analysis and documentations as another source of data. We present results from a systematic search that includes an intensive study on sources and ways of information dissemination of flood-relevant publications. We obtain 186 documents that contain information on the sources, pathways, receptors and/or consequences for any of the 40 strongest trans-basin floods in Germany in the period 1952–2002. This study therefore provides the most comprehensive metadata collection of flood documentations for the considered geographical space and period. A total of 87.5 % of all events have been documented, and especially the most severe floods have received extensive coverage. Only 30 % of the material has been produced in the scientific/academic environment, and the majority of all documents (about 80 %) can be considered grey literature (i.e. literature not controlled by commercial publishers). Therefore, ignoring grey sources in flood research also means ignoring the largest part of knowledge available on single flood events (in Germany). Further,

the results of this study underpin the rapid changes in information dissemination of flood event literature over the last decade. We discuss the options and obstacles of incorporating this data into the knowledge-building process in light of the current technological developments and international, interdisciplinary debates for data curation.

## 1 Introduction

Sophisticated methods have been developed and become standard in analysing extremes in time series, i.e. in estimating the frequency and magnitude of natural events. However, different process types hamper the assumptions of the classical frequency analysis. For the field of flood research, Merz and Blöschl (2008a, b) have called for “a shift away from solving the estimation problem to hydrological understanding”. They argue that the existing formal methods for flood frequency statistics need to be accompanied by hydrological reasoning, i.e. need to reflect the hydrological processes. They specifically argue that the hydrological knowledge gained in the past century is often unduly respected and highlight how the systematic combination of a maximum of relevant information from different complementary sources can help to adjust quantitative estimates from formal methods. Likewise, recently, several international and interdisciplinary groups (International Council for Science, ICSU; International Social Science Council, ISSC; and the UN International Strategy for Disaster Risk Reduction, UNISDR) stated that the considerable amount of information already available on natural disasters has not been adequately

deployed and brought to effective use (IRDR, 2011; White et al., 2001).

This phenomenon, which can be described as a failure to synthesize existing knowledge, is known in many disciplines and results from the complexities in the knowledge-building process. This process requires an information structure that facilitates the distribution and accessibility of content, that is, data, information and knowledge, over time and distance (Luzi, 2010). Barriers in sharing and unsolved complexities in structuring and managing knowledge have been identified as the major bottlenecks in the knowledge-building process (Borgman, 2011; Haendel et al., 2012; Whitlock et al., 2010). Numerous initiatives and international strategies have been set up recently to promote open science and the creation of structured knowledge, particularly aiming at (research) data curation and sharing. The basic motivation behind these initiatives (that ship under the frameworks of open access, open archives, or altogether open science) is the reflection that research results need to be reproducible and verifiable and that data needs to be available to the careful scrutiny of other scientists in order to ask new questions of extant data, to advance the state of research and innovation and finally to create public access to publicly funded results (Borgman, 2011; Wood et al., 2010; Evans and Reimer, 2009; Marx, 2012; The Scientist, 2012). The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities<sup>1</sup> provided a first landmark in 2003 and has been signed by leading scientific institutions and organizations worldwide. It calls for the “... free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display ...” research results. Built on these principles the European Commission in 2012 released a communication, expecting all results – text and data – funded in Horizon 2020 to be openly accessible (European Commission, 2012).

Debate and innovation in data sharing are led by data-rich disciplines such as biosciences or astrophysics and therefore largely focus on observational or experimental data. However, data can take many forms. The National Academy of Science (NAS) in the US uses the term broadly and inclusively, comprising digital manifestations of literature (incl. text, images and many other media) as well as observational data and other forms of data either generated or compiled by humans or machine (Borgman, 2011). For the purpose of this paper we use the term accordingly. A number of applications have already emerged that consider the multifaceted life of data by building repositories that (1) include all kinds of data, from experimental data to publications (e.g. see the Big-Data movement on biocuration, Howe et al., 2008; BioSharing, Sansone et al., 2012; DRYAD, <http://datadryad.org/>; SIMBAD, <http://simbad.u-strasbg.fr/simbad>; see also Genova, 2007), and (2) that deliberately link all information, for example by showing all data sets on a particular genome and all publications that make references to this genome

(text2genome project, <http://text2genome.smith.man.ac.uk/>; see also Haeussler et al., 2011).

In flood research, barriers in access and exchange of hydrometeorological data hamper the setup of central and openly accessible repositories of observational data (Viglione et al., 2010; Hannah et al., 2011). Existing data sets often lack essential metadata to contextualize and interpret time series (Hannah et al., 2011) often providing little-detailed and no annotated information on location, station or catchment alterations, no specifics on extremes etc. Beside observational data and model outputs, data on past events in the broad NAS-definitions sense are another important source of information on which risk assessment and decision making needs to be based. In that sense, the role of past and current natural hazard events as learning examples has been stressed at many instances (e.g., Hübl et al., 2002; IRDR, 2011). Knowledge on flood-specific occurrences is particularly important when interpreting extreme value statistics of long time series, particularly when attributing trends to some causal mechanism (see Merz et al., 2012, for a critical treatment on the current state of trend attribution) and for understanding differences in disaster consequences.

Data on past flood events and in particular event documentations are largely non-research data. Since flood risk assessment is foremost a subject of high societal relevance, it is inherently a subject of governmental action. A large body of authorities is concerned with the management of this risk and the planning of measures for flood loss reduction. Authorities are the primary body of (observational) data and information production and can claim to hold a high level of long-term (technical) experience. They are responsible for maintaining the national station network and are therefore equipped with first hand access to and control of the quality of the data, including data normally not available to the scientific community (different parameters, higher spatial and temporal resolution). Additionally, engineering knowledge, i.e. knowledge on (defence) structures, changes to these structures and their operation both in normal times and during events can mostly only be found at the responsible authorities or operators. Consequently, authorities rather than the scientific community are involved in the production of reconnaissance reports in the aftermath of a flood event. These reports often not only address the hazard part but also provide a more holistic and possibly more detailed view on the event including sources, pathways, receptors and/or consequences.

The mostly technical documents produced by these authorities are commonly disseminated through other means than the scientific publication routes. In information science they are referred to as grey literature. Grey literature is defined by the Luxembourg Convention on Grey Literature (Farace and Schöpfel, 2010) as “that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers” and, as opposed to “white” or “conventional” literature (books, scholarly journals etc.), “where

<sup>1</sup>[http://oa.mpg.de/files/2010/04/berlin\\_declaration.pdf](http://oa.mpg.de/files/2010/04/berlin_declaration.pdf).

publishing is not the primary activity of the producing body". Mackenzie Owen (1997) further highlights that the term grey does not imply any statement on the quality of the document. Rather, it is a characterization of the distribution mode.

In the scientific community, grey literature seems to be largely ignored in the knowledge-building process. Anecdotal evidence and everyday experience confirm two main reasons that have also been found by studies investigating the use and influence of grey in science and research synthesis (see MacDonald et al., 2010, and Rothstein and Hopewell, 2009). First, "white" literature in the form of journal articles or textbooks is trusted since it is perceived as quality-labelled according to its production process that includes peer-reviewing and editorial control. Second, and probably most importantly, practical aspects of information retrieval hamper the use of grey in science. White literature is easily searched for and found since it is at the interest of the (commercial) publishers to make their products available, therefore ensuring both the bibliographic control of each item and provision of an information structure that facilitates the distribution and accessibility of the content (through, e.g., Scopus or Web of Knowledge). To the contrary, accessing grey requires considerably more effort. Sources are dispersed amongst a multitude of producers or custodians that spend less effort in making the electronic metadata or the full text of the document accessible (see e.g. Auger, 1998; Farace and Schöpfel, 2010; Ranger, 2004, for a detailed analysis). Grey documents are mostly produced in the national language of the producing body, making it difficult to find and understand the content for a non-native speaker. However, grey can provide a significant added value to journal publishing through the considerably greater detail at which a topic can be treated in a report and through the content of unique and significant scientific and technical information that is often not included in scientific journal articles or that otherwise is not published at all (Ranger, 2004; Weintraub, 2000; Farace and Schöpfel, 2010).

So far, no systematic approach has been presented that would allow one to defer the size of the body of publications relevant for an event-based assessment of floods, and no reliable estimate can be made on its potential for combining existing knowledge like that contained in flood reports with a data-based analysis. This paper's objective is to first of all identify the existing body of literature that is potentially useful for the specific task of understanding trans-basin floods. Trans-basin floods are extreme events occurring on a regional scale and across catchment boundaries. We aim at creating an openly accessible database of the metadata of publications that contain information on the sources, pathways, receptors and/or consequences for any of the top 40 trans-basin flood events in Germany (as presented in Uhlemann et al., 2010) that can be used as another source of data for flood research.

In this study we develop a systematic search approach with a strong focus on grey publications; i.e. we review the administrative and information landscape in Germany to obtain

an overview on the relevant institutions and tools available for the search. Based on the search results we will elucidate the accessibility and origin of the documents and want to capture their basic bibliographic characteristics. Further, we want to analyse the frequency at which trans-basin floods have been reported and want to assess whether and what kind of changes have occurred in the production process during the study period. This will allow us to determine the potential applicability of flood-event-related publications both retrospectively as well as for future flood events. Further we will discuss the technical options on how this knowledge can be best deployed for future flood research.

## 2 Systematic search approach

When aiming to identify the existing body of literature available on trans-basin floods in Germany for the past 50 yr, the search effort has to include both scholarly sources as well as grey sources. In order to draw reliable conclusions from the search this requires a rigorous and transparent search strategy. We apply the analytical steps and rules developed for systematic reviews that provide the methodological rigour needed for the purpose of our study. In particular, we capitalize on the conception of search strategies as they are defined in systematic reviews.

Generally, a systematic review aims at accessing, appraising and synthesizing scientific information (Centre for Evidence-Based Conservation, 2010); i.e. it is applied whenever there is a need to synthesize the available evidence for a given question, to identify and assess consistent findings across diverse studies (i.e. statistical analysis of causal linkages, effectiveness of interventions) and to inform policy (Burton, 2010; Borenstein et al., 2009). It has originated from the medical and health service sector and become a recognized and standard method also in the field of environmental research and management (see e.g., Higgins and Green, 2011; Centre for Evidence-Based Conservation, 2010). In the field of natural hazard research, systematic reviews have recently been added to the set of research methodologies for Forensic Disaster Investigations (FDI) of the IRDR (Integrated Research on Disaster Risk initiative) (Burton, 2010; IRDR, 2011).

According to the guidelines for systematic reviews in environmental management (Centre for Evidence-Based Conservation, 2010) a search strategy is an a priori description of the methodology to be used to locate and identify studies pertinent to a systematic review. Based on a specific task at hand, it includes a list of search terms to be used when searching electronic databases, websites, and reference lists and when engaging with personal contacts, and the formulation of a priori inclusion criteria (eligibility criteria) that are applied to the search results. In order to ensure transparency and reproducibility of the results, the method requires the documentation of the entire search strategy. In this way it

also entails the opportunity to extend and update the search. Our study differs from a full systematic review approach in that no meta-analysis is planned (meta-analyses are quantitative procedures to statistically combine the results of studies; Cooper et al., 2009). Therefore at this point we do not address the points of study characterization and quality assessment as well as data extraction from the relevant studies. However, when aiming at synthesizing information on flood events from multiple sources as well as for comparing events, the quality of the sources of information needs to be carefully evaluated. We address this point in a separate paper and present a generic framework for quality assessment of natural hazard event documentations (Uhlemann et al., 2013).

In order to develop our search strategy and in particular as we aim to include grey sources in the search, we need to review the administrative and information landscape in Germany to obtain an overview on the tools available and relevant institutions for the search.

## 2.1 The German information landscape

Figure 1 depicts, in a generalized way, the various administrative levels that are concerned with flood risk and/or water resource management in Germany. Governmental agencies are organized into federal government agencies and agencies in the single federal states. Within each the organizational levels range from supreme (ministries) to lower administration (district offices). The institutional hierarchies vary in the federal states; Fig. 1 shows the example of the state of Saxony. Further, the relevant cross-federal and cross-national levels of organizations that are largely structured according to shares in river basin are depicted in the figure. It is important to note that Fig. 1 represents the current administrative landscape and relevant organizations and that numerous changes and reforms have altered this landscape over the past decades.

In Germany, a copy of any published document from public administration is obliged to enter into the German National Library (Deutsche Nationalbibliothek, DNB) and into the respective state library in the federal state where it has been produced (statutory copy). However, the degree of acquisition varies from state to state. Catalogues of the supra-regional to national (DNB) libraries are generally publicly accessible. Further they are part of union catalogues which in turn are part of a widely used meta-search portal for Germany, the Karlsruhe Virtual Catalogue (Karlsruher Virtueller Katalog, KVK). For German publications and in particular for grey publications the KVK can be regarded as the standard national search gateway. It allows for simultaneous search in the union catalogues, hence covering the entire German scientific libraries landscape.

Internationally, in an attempt to combat the threat of losing existing knowledge and to reduce the barriers for an effective use of grey literature, over the past two decades major national institutions and libraries have started creating special

“grey” collections. The most important initiative in Europe is the information system OpenGrey (<http://www.opengrey.eu>), hosted by the Institute for Scientific and Technical Information (INIST-CNRS), France.

## 2.2 Search strategy

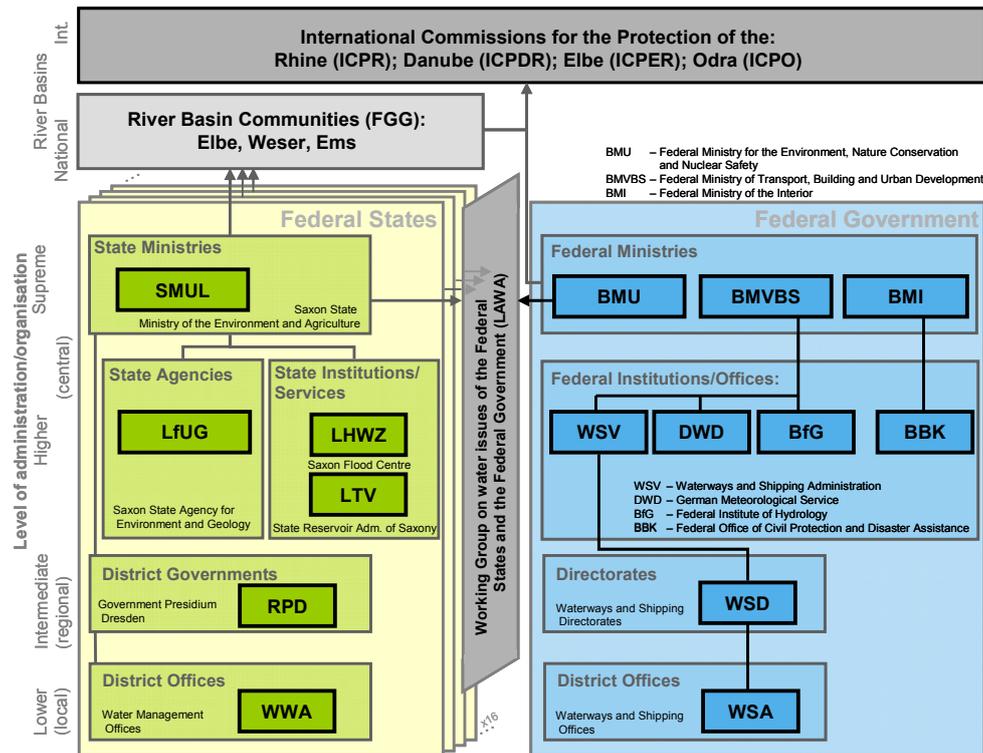
In the following sections we present the single steps of the systematic search for this study according to the methodological design of systematic reviews.

### 2.2.1 Task at hand

The results of any systematic review are strictly related to the chosen question (task at hand) and the search strategy pursued. In our study we aim at identifying flood-relevant literature for the particular purpose of understanding trans-basin floods. Therefore, instead of aiming to conduct a complete search for flood-relevant publications we rather aim for consistency in the search approach. The task at hand for this study is phrased as follows: *Identify all studies that contain information on the sources, pathways, receptors and/or consequences (SPRC) for any of the top 40 flood events contained in the set of trans-basin floods (Uhlemann et al., 2010) and concerning the territory of Germany.*

This task at hand and the expectation of consistency thus impose several logical constraints to the choice of search terms, the search strategy and the formulation of the inclusion criteria. Consistency needs to be given in three particular aspects:

- Temporal and contextual consistency: the strongest constraint results from the limitation of the search to a selection of flood events (top 40 trans-basin floods) and to information on their respective sources, pathways, receptors and/or consequences.
- Scale and spatial consistency: the scale of the publications needs to match the spatial scale and extent of the flood event (large-scale flooding, nationally confined to Germany). The scale consistency has implications on the choice of search tools, the languages for conducting the search and the types of references. In particular, the search can only be consistent (in Germany) at the level of white literature and publicly accessible grey literature of the higher governmental administration and national or international institutions as outlined in Fig. 1. At this level, commonly used and publicly available tools for searching literature are available. Below higher administration the search volume inflates tremendously as the number of relevant administrative units that would need to be addressed inflates, and the search mode would likely need to be extended into archival search. Under the limitations of the resources of our study this approach is not feasible as it could not be conducted nationally coherently.



**Fig. 1.** General organization plan of water management and related authorities in Germany in 2011, i.e. those relevant for floods. The number of levels at each federal state may vary. Also, numerous reforms have altered the organizational structure during the 20th century (ongoing process!). For the federal states the administrative bodies are given as examples for the state of Saxony (dating to 2002).

- Accessibility consistency: with our study we address the scientific community. In order to be consistent with the daily scientific search routines, the search tools for this study need to be readily available also to any other researcher and at adequate expense. Further, any material that has restrictions on public access per se (like confidential material) will not be considered in our study.

### 2.2.2 Search terms and tools

Table 1 provides an overview of all search terms that are used in the English and in the German searches. Any search is limited to the title field. Two sets of search terms are used (flood/inundation terms and defining terms), with individual terms separated by Boolean “OR” operators and sets combined using “AND”. Wildcard symbols (indicated by an “\*”) are used where appropriate. The particularities in German grammar and spelling (compound words and word conjugations) require forward and backward truncation and wildcard replacement of special characters to ensure full coverage of all titles. Generally, any search is the combination of one flood term with at least one of the defining terms.

In Sect. 2.1 we provide an overview of the institutional landscape of governmental authorities on the various levels of administration (Fig. 1). As outlined, in order to conduct a

consistent search we limit the systematic search to organizational levels at river basin scale and the supreme and higher governmental levels to reduce the degrees of freedom in the search and to match the spatial scale of trans-basin floods with that of the administrative levels. Regional and district levels are not approached during the internet search or expert contacts. Further, only scientific libraries at the level above local are included in the search. The Karlsruhe Virtual Catalogue (KVK) is limited to searching only national catalogues, WorldCat and Amazon books. No archive search and no media analyses (web news, print news etc.) are performed. Using the set of predefined search terms, the following search order and respective tools are pursued for the strategic search:

1. Scientific bibliographic databases: Web of Knowledge (in two versions: v. (a) covering the entire period from 1896 to date but excluding conference proceedings, Science Citation Index, SCI; and v. (b) limited to articles past 1990 but including conference proceedings, SCI plus Conference Proceedings Citation Index, CPCI).
2. Meta-search of public open access library catalogues through KVK, and OpenGrey.
3. Search in catalogues not included in KVK: libraries and/or experts of federal or state agencies (e.g. the

**Table 1.** List of search terms.

Sets	Flood (SPRC) terms	Defining terms		
		Year <sup>a</sup>	Place <sup>b</sup>	Month/Season
English	flood*, precipitat*, damage*, event	Tb-flood years between 1951 and 2002; historic*, century	Rhein/Rhine, Neckar, Main, Mosel/Moselle, Donau/Danube, Ems, Weser, Elbe, Mulde, Saale, Oder/Odra Bayern/Bavaria, ..., Sachsen/Saxony	April ... December spring, summer, autumn/fall, winter Christmas, Easter, "new year*"
German	*hochwasser*, *berschwemmung*, *flut*, *regen*, *niederschlag*, *schaden*, *ereignis*	Tb-flood years between 1951 and 2002; historisch*, jahrhundert*	Deutsch*/German*, europ* Schwarzwald*, Alp*, "Bayrisch* wald", Harz*, Erzgebirg*	Januar* ... Dezember*, Frühjahr*, Sommer*, Herbst*, Winter*, Weihnacht*, Oster*, Pfgngst*, Neujahr*

<sup>a</sup> Where appropriate the year is searched for by using 19\* or 20\*. <sup>b</sup> For the English search both the English and the German words for places are used.

library of the Federal Institute of Hydrology, BfG); personal contact; search in index lists of technical (German) journals.

- Internet search: Homepages of respective administrations and associations, unions, etc.; Google/Google Scholar searches are only applied to results from the previous search steps to check for full-text access to the document; if the full text is not available the document is ordered via interlibrary loan.

This is a cumulative process and per iteration only those documents are added to the results list that had previously not been found. All relevant results from the strategic search are included in a reference database using appropriate reference management software.

### 2.2.3 A priori inclusion criteria

Using the set of predefined search terms, the results are reduced to fit the task at hand. The inclusion criteria are applied to the title of each document and, where available, to the abstract provided. Abstracts are commonly only provided for documents listed in the SCI; for any of the meta-search tools for library catalogues only the bibliographic entries are available. Documents with indistinctive title are attempted to be retrieved and are then checked for inclusion.

According to the consistency criteria, only documents that report on any of the top 40 events of trans-basin floods observed between 1952 and 2002 on the territory of Germany are included in the results list. Included are event specific reports and reports that consider any of the contextual criteria of the source, pathway, receptor, and/or consequence framework for any particular flood event. We deliberately exclude studies on water quality aspects and environmental effects such as soil contamination, effects on species or habitats, or sediment transport. Also (personal) experience reports or

narratives are not included. As most river basins in Germany have significant upstream reaches in other countries it is useful to evaluate search hits also from Austria (Danube), Switzerland (Rhine) and the Czech Republic (Elbe) and to a very limited degree from Poland (Odra) (none of the top 40 floods in the trans-basin flood event set exhibited major flooding at the Odra; compare to Uhlemann et al., 2010). If no additional information on source, pathways, receptors and/or consequences in Germany is obtained from these documents, they are not included in the results list. Further, only reports with a regional scope or broader are included. Local studies that analyse or document the event at a district or city level are not considered. This is sometimes difficult to obtain from the title of a report, and some reports on local aspects also account for the regional aspects of the flood, i.e. in the description of the hydro-meteorological causes. They are then also considered.

We include solely print material (both paper and e-prints). According to the strategic search only website contents of either scientific or agency origin are included. This excludes reports from Wikipedia, newspapers, internet news pages, broadcasting (videos, audios) and social networks. Also material in the form of presentations, mostly power points of meetings, classrooms, conferences, etc., are excluded.

### 2.3 Document characteristics

By analysing the metadata of each document that was retrieved through the systematic search and by classifying the document along event-specific aspects, we aim to identify the key players in report production including a characterization of the production process and want to characterize the material potentially useful to maximize the information per event (who produces what, when, how and why).

In order to identify the main producers of flood-event-related literature, we associate the author(s) or issuing

institutions to 7 classes according to their affiliation: (1) specialized governmental agencies (any of the federal or states level shown in Fig. 1 and commissioned with flood/water management tasks); (2) non-specialized governmental agencies (governmental agencies not particularly commissioned with flood/water management tasks, mostly ministries); (3) non-governmental organizations or associations; (4) intergovernmental/international commissions (e.g. ICPR, see Fig. 1); (5) science/academia (research centres, universities), (6) business (e.g. insurance companies, associations for shipping, etc.); and (7) other or unknown affiliation.

Further, to analyse what is being produced, we classify the reference type of each document. Table 2 lists all classes that are accounted for.

To analyse the accessibility of the material, we analyse the results of the strategic search with respect to how the document was found (level of search: SCI, KVK, OpenGrey, homepages, reference lists, etc.) and evaluate the extent to which the document's full text is openly accessible and which format (electronic or print) the documents have upon retrieval.

We provide a report typology that basically classifies the purpose of the document in terms of its specificity in being related to any particular flood event. Table 3 lists the classes and the definitions of each class.

Irrespective of the report typology assigned, each document contains information on one or many trans-basin flood events. For each document the full list of events is recorded, including the month and year of the flood and the rank given to the flood in the set of trans-basin floods, allowing linking each document to the set of trans-basin floods and the existent characteristics per event. Also it allows identifying the number (and types of documents) available per event.

### 3 Results and discussion

The entire database of all references and the full evaluation table of the document characteristics are accessible via the Supplement that can be accessed using the doi provided in Uhlemann (2012).

#### 3.1 Systematic search

Using the set of predefined search terms, the systematic search resulted in the identification and acquisition of 186 documents that fulfilled the inclusion criteria. In the cumulative process of the search the metadata of an initial set of 26 documents was identified using the Web of Knowledge. The largest share of documents (114) was identified using the KVK (excluding documents that can be found through SCI). Using the OpenGrey platform resulted in generally very few relevant hits, none of which was additional to the searches performed on KVK or SCI. Additional material was then found through searches at institutional homepages (13),

**Table 2.** Classification of the reference type.

Reference Type	Definition
International Scientific Journal (CI)	Article in a peer-reviewed, SCI-listed journal
(National) Scientific/Technical Journal	Article in a national journal, not necessarily peer-reviewed
Specialized Periodical (reg.)	Regularly published, mostly governmental continuous reports (e.g. hydrologic year books, monthly weather reports etc.)
Technical Report/Series (irreg.)	Technical Report or report published within a technical but irregularly appearing series
Monograph	Books, thesis
Edited Material	Article published in an edited book, a book section or as part of conference proceedings
Expert opinion	Commissioned work not produced directly by gov. agencies
Web page	Material that is published only online (i.e. on special portals)
Brochure	Material published for information of the public
Press Release	Releases from official authorities to the press
Other/Unknown	Document does not fit any of the above classes or type of reference can not be identified

through checking reference lists (15) and from tables of content provided for some technical non-SCI journals (4). Fourteen (14) sources that were otherwise not found were obtained from the special library of the Federal Institute for Hydrology (BfG).

The effort in accessing the material varied considerably with the main differences originating from language specificities of the document and the technical capabilities of the search portal.

The SCI provides comfortable and standardized functionalities in the search options and output formats, including interfaces to referencing software. Publications included in SCI are provided with a link to the abstract of the document (if any is provided) allowing for application of a priori inclusion criteria directly on the search results. This allows for an efficient search as the document does not need to be acquired before the inclusion criteria can be applied.

Search results conducted through the KVK cannot be saved or exported as no interfaces are provided. Lists of results are provided separately for each of the included union catalogues, leading to highly redundant search results. Also, keywords and abstracts are not provided along with the metadata. Therefore, if the title of the document is inconclusive with respect to the a priori criteria for inclusion of the document in the search results, the publication first needs to be acquired, which is challenging with respect to financial and time resources. Further, at the time of this study the technical capabilities of the portal were partly limited with respect to the error-free transmission of search terms to the embedded union catalogues. This meant that the particularities in

**Table 3.** Report typology

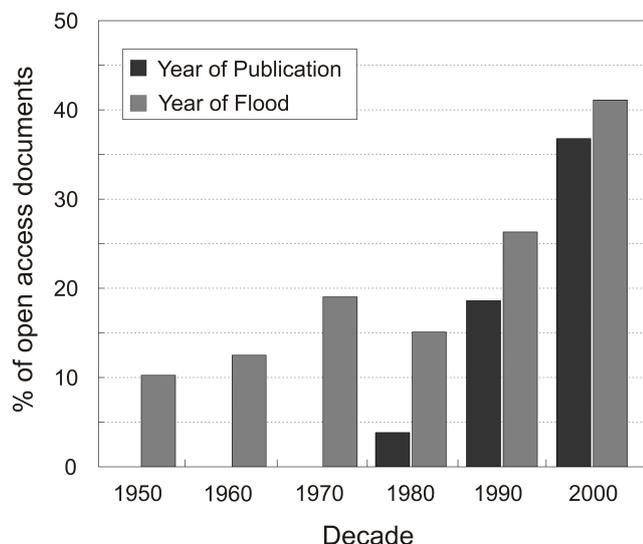
Report Type	Definition
Special Report 1	Report on one or possibly two particular flood events aiming at documentation and analysis. If two events are treated, they are described together due to their close temporal occurrences and/or related causes.
Special Report 2	Reports on two to five (rarely more) events, sometimes with the aim of comparative analysis but generally aiming at an event description.
Special Report 3	Reports or studies on certain aspects of flood analysis making reference to case studies (i.e. any trans-basin flood). Lessons learned studies (any aspect).
Regional Report	Reports with a regional perspective (geographical region or particular river/basin) either presenting (extreme) flood event collections or studies on flood characteristics in that region that also contain useful information on a particular event.
Continuous Report	Official documents issued by governmental authorities for the purpose of data publication and continuous documentations of, e.g., the state of rivers, water resources, etc. In case of hydrologic yearbooks or monthly/quarterly continuous reports, flood events are naturally included. For meteorology also the effects of hydrometeorological events are listed (not consistently but frequently).
Other	Reports fitting none of the above classes.

German grammar and spelling (compound words and word conjugations) that require forward and backward truncation and wildcard replacement of special characters needed to be partially substituted by full-length word searches (see combinations of search terms of Table 1), therefore inflating the search.

Searches conducted directly at the producing body or their associated libraries (that are not included in KVK) also proved to be less straightforward as complete lists of all publications and the provision of central access points or search options to a database of publications are not the rule. We find that except for the Federal Institute of Hydrology (BfG, Bundesanstalt für Gewässerkunde, 2012) no authority has provided an overview on its entire list of publications. The institutional library of BfG maintains a very large collection of flood-relevant literature; however, the stock is largely confined to material concerning western Germany, the catalogue is not part of the meta-search portal KVK and old material is largely not searchable as it has not been included in the digital catalogue (the largest share of old publications (< 1980) has not been entered). However, in the course of the study we obtained copies of paper records archived on microfiche that allowed us to also search the catalogue for documents published prior to 1980. Many authorities provide a publication list on their homepages; however, they are not complete: they mostly do not list publications before 1990 nor do they include scientific/technical articles submitted to journals by individual employees or as results of cooperation. Further, following up on the discussion on open access to digital works provided in the introduction of this study, a decade

later we have to subscribe to the findings of Warnick (2001) that no agency has systematically digitized its legacy collection. Recent publications are frequently added as digital and downloadable documents on the authority's web pages; however, this has not resulted in the automatic indexing of the document's metadata in an electronic database. In as much as this improves the access to full text, if the users' search strategy solely relies on searching electronic databases, then these documents will not be found.

Access to the electronic full text of any of the identified documents depends on the number of journal subscriptions and interlibrary loan agreements that are provided by the hosting institution within which the search is conducted. Given the licenses and subscriptions at hand for our study, 49.2% of all documents were retrieved as print material. Electronic, machine-readable text was obtained in 36.3% (33.5% as pdf, 2.8% as online material) of all cases. Electronic but not text-processable scanned documents in pdf form comprise 14.5%. In sum, 22.7% of the documents identified for this study are fully openly accessible, most being provided on agency web pages. Figure 2 shows the percentage of documents that are openly accessible (OA) per decade. For one, the percentage is given for the number of documents produced on floods in a particular decade, and, second, the percentage is given for the actual years in which the documents were published. The figure highlights that the share of publications with OA increases strongly with time. In particular it highlights that only reports published past 1980 are OA and that OA reports on events prior to 1980 have actually been published past 1980.



**Fig. 2.** Percentage of documents with open access to the full text, aggregated into decades (2000 only accounting for 3 yr, 2000–2002) considering (a) the year when the document was published and (b) the flood event years.

### 3.1.1 A note on completeness of the search

In our study we conduct a systematic search for publications relevant for the task at hand (trans-basin floods) and our search strategy builds on three criteria of consistency: temporal and contextual consistency (documents that contain information on sources, pathways, and/or receptors and consequences on any of the top 40 trans-basin floods), scale and spatial consistency (conducting the search only at the organizational levels that correspond to the large-scale type of the floods and the respective search tools available for that purpose), and accessibility consistency (limiting the search to material that is potentially accessible to any researcher). Our study therefore provides an estimate of the size of the body of literature that is specific for the task at hand. Completeness then has to be assessed with respect to whether the search strategy is comprehensive enough to be able to identify all material actually accessible under the given search scope.

At the level of the search that is independent from the producers of documents, our search is exhaustive as we deploy the entire spectrum of search tools available to retrieve scientifically indexed publications as well as grey publications indexed in online public access catalogues (OPAC) in Germany. Misses at this level of the search may result from the technical limitations of the KVK, i.e. the partially dysfunctional transmission of search terms to the embedded union catalogues. However, we estimate that this concerns very few documents as most documents have been indexed in more than one library catalogue and can therefore still be identified. Misses may further occur due to lagging inclusion of older paper records into any OPAC. However, since we could

access the full catalogue (paper and electronic catalogue) of the BfG library, which serves as a national collection for any kind of hydrologic publication, we also estimate that the number of missed documents is rather small.

At the level of the search that addresses the producing bodies directly, we include the homepages of all higher and supreme agencies of the federal and state government, including that of international organizations. This search addresses the present form of the administrative organization, and the search results depend on the degree to which documents of previous structural units have been included in the current homepages. In any case, the statutory copy obligation has been in place for the entire period investigated in our study (in western Germany). Therefore we estimate that the effect of missing documents from former governmental units is more a matter of the afore-mentioned lagging inclusion rates of file cards in OPACs.

Beside the accessible material it is likely that more grey literature has been produced that might contain information on the SPRC of trans-basin floods in Germany in the considered period but that has not entered the electronic databases so far or that falls outside of the scope of the search (i.e. does not meet the consistency criteria). The size of this body of literature cannot be inferred at this point of the study as it would require a different search strategy and a substantially larger effort. Based on the experience gained during the search in our study, we conclude that an (expert) survey amongst the producing bodies and custodians of knowledge would be necessary in order to obtain a best estimate of the material produced at (all) governmental institutions. Also, by extending the search to archives and non-public catalogues (governmental and institutional) a number of additional documents might be detected. This concerns particularly older documents (not included in an OPAC so far), documents produced on the lower organizational level of public administration and those of the former German Democratic Republic. For the latter it is, however, not unlikely that due to the restrictive and centralized publishing strategy in the GDR (S. Kühnert, personal communication, 2012) the number of detectable documents remains small. In any case, archival search requires adequate skills and is extremely time consuming, and the cost–benefit would need to be carefully evaluated.

### 3.2 Basic characteristics of the material

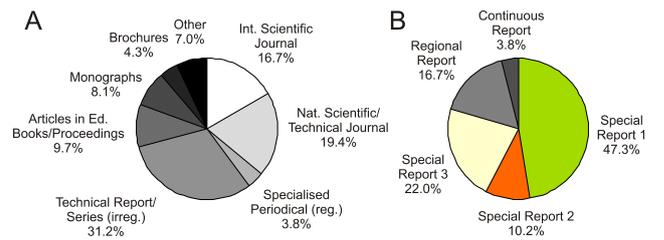
We start analysing the material based on the metadata characteristics described in Sect. 2.3. Each document is referred to once; double counts due to multiple events described within one document are not considered at this stage. The study of the material revealed that two out of the first 40 events in the set of trans-basin floods are generally documented in pairs treating two consecutive events as mutually dependent. Both events are merged with the respective dependent events for the purpose of this study (December 1993/January 1994

flood, ranks 8 and 19; February/March 1999 flood, ranks 25 and 27). The strategic search is extended for two more ranks (to 42).

Figure 3a displays the reference types that can be ascribed to each document. Nearly one third of all documents belong to the group of technical reports or reports that are produced in irregular series. A total of 17 % of the relevant material has been published in international SCI-listed journals, 19 % in technical journals that are published mostly for the national market, often specialized for particular regions or branches. A total of 18 % of the material is comprised of either monographs (including books and theses) or articles in edited books or conference proceedings. Specialized regular periodicals, i.e. yearbooks, monthly reports, etc., contributed by 4 %. The remaining material is evenly spread amongst brochures and other material (expert opinions, web pages, press releases, etc.). Based on the definition of grey literature (Luxembourg Convention) the material is up to about 80 % comprised of grey items, and only the SCI-listed journals as well as part of the monographs and edited books can be considered as fully white literature.

Using the typology of reports presented earlier we analyse the specificity of each document with respect to a particular flood event; see Fig. 3b. Over half of the material is comprised of reports that were specifically produced to document or analyse one particular flood event (Special Reports Type 1, 47 %) or several events (Special Reports Type 2, 10 %). Reports of a mostly scientific nature, investigating certain aspects of floods and making reference to a case study (any trans-basin flood) or lessons learned studies, contribute 22 % of the material; regional studies make up 17 %. For the moment, the share of continuous reports in the set of material only forms some 4 %. We only count each type of continuous report, not including each of the issues produced. The systematic search revealed that yearbooks and monthly reports on both hydrological and meteorological aspects have been produced for nearly the whole period and both on national and regional scales. However, in the course of this study only a limited number of issues from the continuous series could be retrieved. Acquiring all this material will significantly change the share of continuous reports, and it will be interesting to systematically analyse the information content of this material in future.

Most of the material is produced within the national context and for a national auditorium. Only 7 % of all documents have a cross-border or European scope. A closer look reveals that European-scale analysis can be found only for the most recent and also (probably) most damaging flood event of August 2002. This flood affected the central European space. The remaining material has in 27 % of all cases a national scope and in 61 % of the cases a regional (often related to a particular basin or river) or federal scope (due to the federal jurisdictions). According to the search strategy that largely excluded local searches, the share of material that has a very narrow spatial focus is small (5 %). Further, as a consequence



**Fig. 3.** Typology of all documents by (A) reference class and (B) with respect to event specificity.

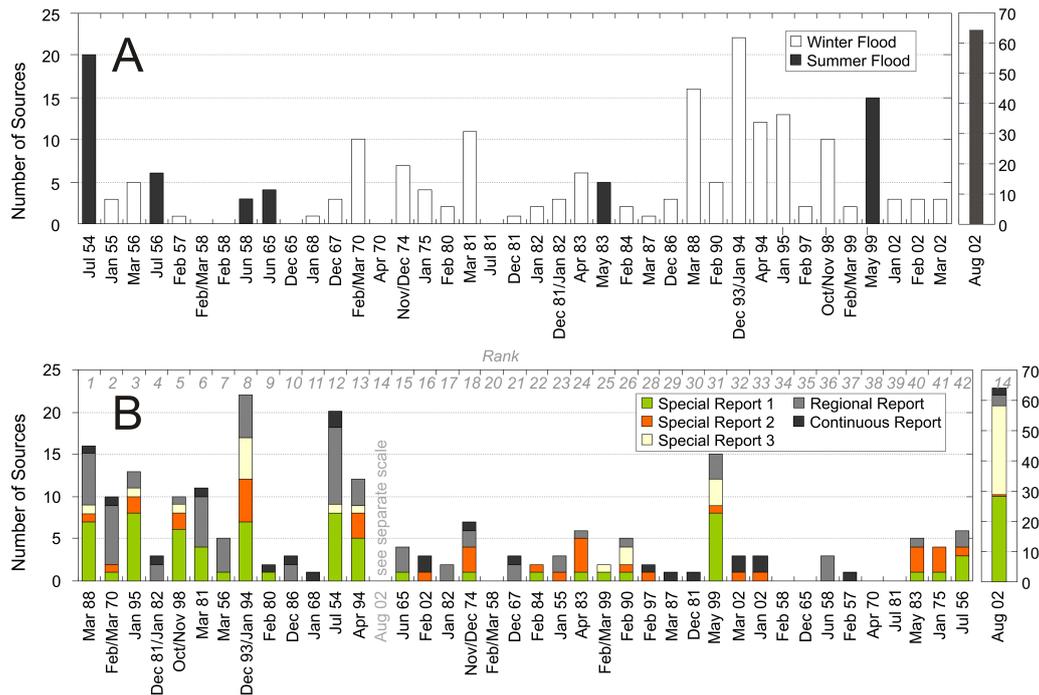
of the national task at hand, the main language (text body) of the retrieved documents is German (81 % of all reports); 18 % are published in English, and less than 1 % in other languages.

Using the classification for the affiliation of authors and/or producing bodies, the analysis of the material reveals that the majority (54 %) of documents retrieved was produced by governmental agencies (or their employees) that are specialized in the field of flood or water management. A total of 30 % of all relevant documents were produced in the scientific/academic environment and 3 % by intergovernmental commissions. The remaining percentages are contributed as follows: 5 % by higher level, non-specialized governmental institutions (mostly ministries), 2 % by non-governmental organizations, and 4 % by business; 6 % of the authors could not be associated with any particular institution. From the 54 % of documents produced by specialized governmental agencies the largest share was produced on the states level (66 %), and there nearly exclusively by the state agencies. At the national level, agencies associated with the Ministry for Transport, Building and Urban Development (BMVBS; see Fig. 1 for the overview) contributed to 25 % in total, with 14 % by the Federal Institute of Hydrology (BfG) and almost all publications main-authored by one person (H. Engel), 7 % by the German Meteorological Service (DWD) and about 4 % attributable to the Federal Waterways and Shipping Administration (WSV) (or its subdivisions). A total of 7 % of all documents were produced on a national level in the former GDR.

### 3.3 Event coverage

On an event basis, the analysis crosses over the entire set of documents per flood event and then summarizes over all events. The amount of reports for this analysis increases from 186 to 272 as documents that contain information on more than one event are listed several times.

The amount of documents that contain relevant information on any of the 40 trans-basin flood events varies considerably. Figure 4a shows in chronological order the total number of documents per event. For 5 out of the 40 events (12.5 %) no relevant document could be retrieved. For the majority of events (60 %) less than 5 reports were found. A total of 8



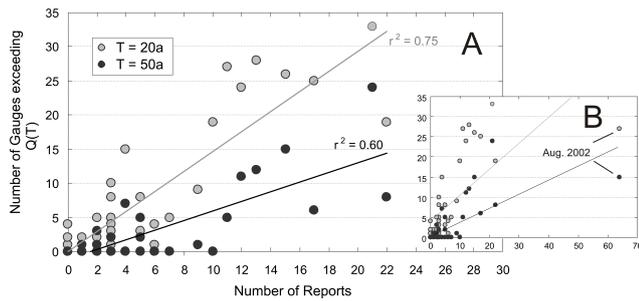
**Fig. 4.** Number of documents per event for the top 40 trans-basin floods in Germany. **(A)** In chronological order; **(B)** in order of trans-basin flood severity and stratified by report typology. The August flood of 2002 is displayed at a separate scale.

events were documented by 5 to 10 reports, and 20% of all events received extensive coverage of more than 10 reports – in three cases, of more than 20 reports. Figure 4a also denotes the season during which the flood occurred (stratified in summer and winter half years: May–October and November–April, respectively). Both floods that received the highest reported frequency (July 1954, August 2002) are summer floods. Generally, considering special reports of type 1, summer floods tend to be more intensively reported on. This effect can be largely attributed to the level of damage that was encountered during the flood. Floods with a large spatial extent but with less-severe local magnitudes and consequently less damage often do not draw major attention. Many of the winter floods are characterized by this phenomenon. Trans-basin summer floods are characterized by high local magnitude, often with record-breaking rainfall intensities, flash-flood characteristics in headwater catchments and high damages including fatalities.

Uhlemann et al. (2010) provide an index of severity that allows comparing events according to their spatial extent and their patterns of magnitude. Using this index, Fig. 4b depicts the events in their order of severity, revealing that generally the highest-ranking events are also those that have been reported on most. Considering the amount of special reports type 1 and 2 produced per event, this effect is even more pronounced. The seasonal effect described earlier leads to summer floods generally receiving more attention than winter floods of comparable rank.

The influence of flood magnitude on the number of publications per event is illustrated in Fig. 5, where magnitude is expressed as the exceedance of a certain return period at a certain number of stations (using the 162 gauges of Uhlemann et al., 2010). If return periods were encountered that equalled or exceeded the 50-yr flood, almost certainly at least one report was produced and made accessible, and at a level of the 20-yr flood the publication of a report is very likely. For both return periods a linear regression can be drawn (Fig. 5a) and the correlation coefficient is significantly positive with  $r^2$  of 0.75 for  $T = 20a$  and  $r^2 = 0.60$  for  $T = 50a$  (excluding the flood of 2002 from the analysis). Generally, the more gauges exceeded peak discharges above  $T = 20a$ , the more reports were produced. Clearly, damage is the unifying criteria both for report production and also for information dissemination. We find that publications on low-frequency high-damaging events are more likely produced as special reports (SR1–3) and therefore entail a meaningful title (mostly the flood event is mentioned in the title) and, due to public and political interests, more outreach activities are pursued by the producing bodies. In turn, high-frequency but low-damaging floods are more likely to be treated in context (regional reports, continuous reports) with less meaningful titles and with less effort in creating access to the document.

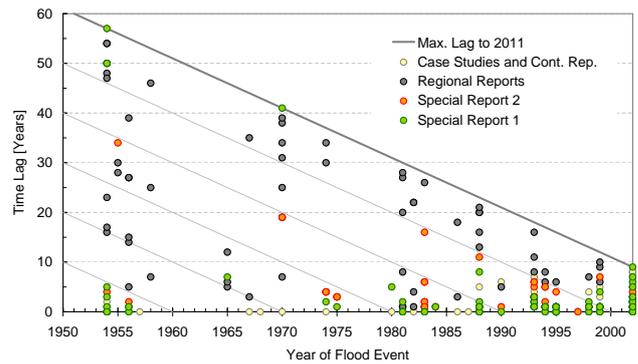
However, the linear regression of Fig. 5a can only roughly explain the threshold behaviour of public interest in a flood event and holds only if the flood of August 2002 is excluded from the regression. Figure 5b shows the same analysis



**Fig. 5.** Correlation of number of documents per event and event magnitude (expressed as number of gauges that exceeded the return period  $T$  (20 yr or 50 yr) during the event). **(A)** Excluding the August flood of 2002; **(B)** including the August flood of 2002.

considering also the flood event of 2002. Given the event magnitude and spatial extent of the flood, the number of reports produced for this particular flood far exceeds the reporting numbers of all previous events. Obviously some other mechanisms than the previous publication strategies were effective for this event. In the following we analyse the changes in publications frequencies over the investigated period.

As shown in Fig. 4a, within the body of material at hand for the analysis, the number of accessible publications and particularly the number of special reports increases with time. Starting from the late 1980s this is very apparent as nearly each event starts to receive special documentation (SR1 or SR2). Figure 6 displays the time lag at which a report had been produced relative to the year in which the flood event that the document refers to had actually occurred. For a better orientation Fig. 6 includes a line indicating the maximum possible time lag that a document could have (referring to the year 2011, in which the systematic search was conducted for this paper). Considering the year of publication our results show that 35 % of the material was published within the last decade and 25 % in the 1990s. For the decades of the 1960s to 1980s this proportion is considerably smaller, and it can be seen that only few documents were retrieved in the course of this study that refer back to earlier events. The figure also shows that, whenever special reports are produced, they are usually published immediately. Special reports of type 1 have up to 57 % of the time been published in the year of the flood and 87 % within the first two years. For the flood event of 1954 two reports had been published more than 50 yr later, invoked by a similar flood event (August 2002). Documents other than special reports, i.e. regional reports, are published also a long time after the event. Often these documents implicitly contain information on a particular event of the past, for example in the analyses of the flood history of a region. The likelihood of a flood event being included in a regional analysis naturally increases with time, and consequently the share of regional reports in the number of publications produced per event is larger for floods of the past.



**Fig. 6.** Time lag between the year in which the document was published and the year of the flood event. For better orientation the diagonals depict decadal isolines referring to the publication year (2011, dark line; 2000, 1990, 1980, 1970, 1960, light grey).

The increase in numbers of documents since the 1980s coincides with three distinct social, political and environmental shifts: (1) the start of the digital period, (2) the reunification of the two German states in 1990 and (3) the onset of a flood-rich period starting from the 1980s (compare to Uhlemann et al., 2010) including many of the strongest trans-basin floods. Due to the clustering of floods in this period, the total number of documents in that period is also largest. However, the number of reports per event has also increased. It can be debated whether this increase results from an increase in reporting frequencies in the last two decades or indicates that documents prior to 1990s are less accessible. We find that both aspects contribute to the effect. In Sect. 3.1.1 we highlight several reasons that may be responsible for a reduced accessibility to older documents. However, the strong increase in publications over time and in particular over the last two decades far outweighs the more limited access to older documents. The clustering and the occurrence of some key events with particularly high damages together with the question of climate change impact on extremes has drawn marked media attention, increased awareness and fear in society and subsequently has led to political action like the creation of (international) river commissions (ICP Rhine founded in 1950, ICP Danube River in 1998, ICP Elbe River in 1990, ICP Odra 1999), the initiation of flood management plans and political frameworks like the EC Water Framework Directive (2000/60/EC) and the Flood Risk Management Directive (2007/60/EC) that in turn led to the creation of River Basin Communities (see Fig. 1), the foundation of a Federal Office of Civil Protection and Disaster Assistance (BBK) for Germany in 2004, the initiation of research priorities and programmes (United Nations: International Decade for Natural Disaster Reduction (IDNDR, 1990s); Ministry for Education and Research, Germany: German Research Network Natural Disasters; DFNK, 2000–2004). Further, authorities have realized that risk awareness can only be created through information dissemination. Beside event-related reports this

becomes clear as the number of publications intended for a general audience in the form of brochures and pamphlets has significantly increased recently. Beside the increase in numbers of publishing bodies on the river scale, it also has to be kept in mind that the number of federal authorities has increased by 6 in Germany as a result of the reunification of the two German states in 1990 and the re-creation of the federal structures in former East Germany. Consequently the restrictive and centralized publishing strategy of former East Germany (S. Kühnert, personal communication, 2012) has been replaced by federal authorities with responsibilities in free and impartial information for the public.

Several studies on flood damages (e.g., Petrow et al., 2006; Thieken et al., 2007) have shown that experience is a large factor in creating a sense of risk awareness, leading also to reductions in damages. The most prominent examples in Germany are the two consecutive flood events in the middle to lower Rhine in December 1993/January 1994 and in January 1995. Our strategic search results show an increase in special reports (Type 1) for the latter event which can be attributed to the high level of awareness still present both in public and administration. In total, the number of relevant publications is larger for the 1993 flood as comparisons are frequent in 1995 flood reports to 1993, hence covering both events. Another effect can be observed as a consequence of these two events: the triggering of scientific event-specific publishing (in the form of SCI-journal or proceedings' articles, books, project reports). Prior to 1993, scientific event-based studies of floods were hardly published, but with the series of flood events in the Rhine region (starting already before 1993 with floods in 1988 and 1990) the topic was put on the research agenda and remained there as further extreme floods occurred within critical periods of time. The community of researchers and projects has since increased, leading to an increased publication output.

The most remarkable effect of scientific contributions has been the tremendous amount of scientific publications on the most recent flood (in the set of events) of August 2002. This event received an exceptional number of publications, showing a distinct difference to all previous events (as shown in Fig. 4). For this event the share of scientific articles is almost equally as high as the number of all other publications. The 2002 event seems to have created an unparalleled type of case study for a wide array of research fields (social, health, economics, risk, engineering, hydro-meteorology, ecology, etc.). Beside this scientific interest also the amount of publications from the agencies far exceeds that of any previous flood and is remarkably larger than for floods of similar characteristics. For example, only half as many documents were found for the flood of 1954, although this event exhibited partially even higher magnitudes in the same region with higher losses. Further, more than 40 % of the documents on this event are openly accessible (compare to Fig. 2), highlighting a distinct shift in the dissemination strategies pursued especially by the agencies.

#### 4 Conclusions

In our study for the first time we present a systematic approach that allows deferring the size of the body of publications relevant for an event-based assessment of floods. That is, this paper's objective was to first of all identify and characterize the existing body of material that is potentially useful for the task of understanding trans-basin floods in Germany from a scientific perspective. Based on the methodological steps developed for systematic reviews, we present a search strategy that explicitly includes grey literature, using the tools widely available to conduct a search in the German information landscape. The search results are reflective of the material that was actually accessible at the time of the search and under the search strategy pursued. The strength of the approach lies in the full documentation of the entire search process and in it allowing the search to be extended and updated.

We obtain 186 documents that contain information on the sources, pathways, receptors and/or consequences for any of the 40 strongest trans-basin floods in Germany in the period 1952–2002. Most of these 40 flood events have been documented (87.5 %) and especially the most severe floods have received extensive coverage. Only 30 % of the material has been produced in the scientific/academic environment, and the majority of all documents (about 80 %) can be considered grey literature. Therefore, our study reveals that ignoring grey sources in flood research also means ignoring the largest part of knowledge available on single flood events (in Germany).

We present the results of our study in an openly accessible database (Uhlemann, 2012). This allows any potential user to circumvent the tedious work of searching for material that is otherwise scattered amongst a multitude of producing bodies and information providers. The presented collection of material can be considered as a repository containing the digital, centralized metadata of documents relevant for trans-basin flood event analysis. In that way, the results of this study are the first step into a structured deposition of content, therefore providing access to existing knowledge that would otherwise likely be ignored, and the database can be considered as another source of data for further flood research tasks.

With this study we not only want to provide a database of flood documentations but also create awareness in the research community that grey sources should be an integral part of the knowledge-building process. The way in which this knowledge can effectively be combined with analyses based on observational data and modelling results has to be defined separately. However, a knowledge base that integrates all data available will largely facilitate this process and may be a precondition to a successful combination of all sources.

So far, there is still a long way to go before a knowledge base on floods is created that could be scrutinized for any (new) scientific query or question. However, a number

of critical conclusions can be drawn from the results of this study on the potential applicability of the current material and the next steps to be taken.

Two main barriers in deploying the identified material for research synthesis remain after completion of this study. This is for one the language barrier as 85% of the material is written in German and only a small share is accompanied by an English title and/or captions. This certainly limits the widespread use of the material, and to allow for interpretability these documents would have to undergo some processing, either in the form of direct translations or in the form of content tagging (keywording, header and caption translations, multilingual ontology that includes dictionaries). Second and most important: so far only a rather small share of the identified documents is openly accessible as full text (in total 22%). The remaining material is available only as a digital meta-data set (see Data description), and the full text needs to be acquired from the publishers (journal) or via interlibrary loan. Even then, upon retrieval a substantial share of documents is still not digitized. Open access to scientific results has been discussed mostly for high-ranking journal articles in the last years. In terms of open access publishing, formerly grey literature could now become an integral part of scholarly publication. Quality-assured reports, published by trusted institutions, typical for grey literature, could be fully accessible, leaving all distribution and accessibility problems behind them.

Digital full text documents are a precondition to apply advanced text-processing tools (semantic-/ontology-based text mining) that facilitate the efficiency of any search query and therefore improve largely the speed of information reception including that of machine information processing. For now, any information within the documents can be processed only intellectually. Providing open electronic access not only to current reports but digitizing the legacy of knowledge of flood related institutions would provide highly useful research data to the scientific community. Making such older research data electronically available is always expensive, including costs for annotating, adding relevant metadata to make it useful and, in the end, providing long-term accessibility (Houghton, 2011). Most institutions do not even see a priority to build such a resource. During our study we found no agency that has systematically digitized its legacy collection.

In order to ensure long-term applicability, extending and updating the knowledge base will be of high importance. This accounts both for past flood events (closing the gap to historic hydrology) as well as for future events. Based on our results we expect that adding material to the database on events after 2002 or any future event will likely become easier and that the amount of material will be more abundant, including coverage also of low-impact floods. We could show that the increase in numbers of publications per event over time results from a combination of factors, with accessibility barriers playing a subordinate but not-unimportant role

and changes in publication frequencies due to coincidence with three distinct societal (start of the digital age), political (reunification of the two German states) and environmental shifts (flood-rich period) explaining most of the change.

These shifts led to an increase in the number of publishing bodies and to changes in publication philosophies, i.e. increased output from authorities as strong floods occurred within short intervals and flood risk management prominently entering the administrative and research agenda. The most distinct change occurred with the central European flood of August 2002. In as much as this flood was exceptional in many ways (large spatial, international extent; extreme damages), it also demonstrated the tremendous changes in recognition of floods at the administrative level due to information needs from the public and due to the opportunities of internet publishing. The publishing strategies at the agencies have changed largely towards web-publishing of reports providing access to the full text. However, they are often far from fulfilling basic standards of electronic publications. Solutions have to be found on how to support agencies in their archiving and publishing strategies; i.e. it will be crucial to develop standards for information dissemination (i.e. guidelines for administration on how to assure persistent citability, or how to license content for scientific re-use as open as possible; Creative Commons licences, CC by). As research has started to contribute to event-based analysis largely in the form of case-studies since the 1990s and most prominently since 2002, the amount of digital publications (not necessarily open access) outweighs pure print publications since 2002. Currently, the high-level political pressure on open access to research results and on the value and access to data already alters the ways of scholarly communication. We expect that scientific publishing will continue to become more and more open and that incorporating this (digital and largely annotated) knowledge into the knowledge-building process will become increasingly faster and simpler.

Currently, particularly for grey, we see a gap between the rapidly evolving or already-available technologies for information cropping (text mining, semantic search tools) and access options to digital content (see Marx, 2012, and Rinear and Palmer, 2009, for future visions as well as critical discussions in Van Noorden, 2012, and Borgman, 2011) and an infrastructure that merely supplies the most-basic functionalities for such a development. It can be expected that the fossil design of KVK will soon be replaced by more-advanced search tools. Already, new search platforms have been established (e.g., “Base” <http://www.base-search.net>). In Germany, the “Wissenschaftsrat” recently published a paper on the perspectives of the German information infrastructure, paving the way for sustainable development of advanced tools (Wissenschaftsrat, 2012). However, in addition to the systematic search strategy presented in this paper good web search strategies will be needed.

For the future of knowledge management the roles of science and libraries in data and knowledge curation need to

be critically reflected (as is currently being done in an international debate; see, e.g., Smith, 2011). New standards in information and data literacy need to be developed for data and information curation, and libraries are the natural experts in data management and can provide the institutional level of support to permeate these standards into the local cultures (Haendel et al., 2012) – that is, both to the individual scientist's level as well as the institutional level including that of public administration. Already, technological developments and international projects in fields such as astrophysics and bioscience preordain the path for the future of data curation through the combined deposition and query of all data available on an object of interest. For hydrology, the development of such standards is currently being discussed for observational data archives (Hannah et al., 2011). In order to develop the full potential of the extant data and knowledge, these initiatives should be either complemented by or at least linked to publication databases, and database developments should be thought together and not separately. As demonstrated by the biosciences (OBO Foundry, 2012; GO Consortium, 2012), semantic compatibility and therefore the setup of a (hydro-)ontology will be the basis to ensure interoperability of all data. Clearly, to provide these functionalities for and the access to databases will require a substantial investment (time and funds) as well as the coordination of many different parties and stakeholders. It will probably not be achieved easily. We believe that the results of our study can contribute and help to structure the current debate on knowledge management and curation for flood research.

### Data description

The data used for this publication is freely available as Supplement under the creative commons license and can be permanently accessed following the doi given in Uhlemann (2012).

*Acknowledgements.* We thank the GFZ German Research Centre for Geosciences and the University of Potsdam for their financial support. The main author is grateful for partial funding through the excellence in teaching programme (Junior Teaching Professionals) of the Potsdam Graduate School and a scholarship by the equal opportunities office, both the University of Potsdam. We want to thank the authorities and respective individuals who provided information and access to relevant publications for this study, who assisted in their retrieval and who contributed donations. We dedicate our special thanks to the library of the Federal Institute of Hydrology for their straightforward cooperation and to the staff of the library of the science park Albert Einstein in Potsdam. The numerous dedicated hours of work of the students Florian Betz, Sylvia Wesser, Annika Meiser, Tim Wahrenberg, and Fabian Gehrmann who assisted in the strategic search are gratefully acknowledged.

The service charges for this open access publication have been covered by a research centre of the Helmholtz Association.

Edited by: G. Di Baldassarre

### References

- Auger, C. P.: Information sources in grey literature, Bowker Saur, London, 1998.
- BfG (Bundesanstalt für Gewässerkunde): Publikationen, Koblenz, available at: <http://www.bafg.de> (last access: 17 August 2012), Wissenstransfer/Publikationen, 2012.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R.: Introduction to Meta-Analysis, John Wiley & Sons, Ltd., Chichester, UK, 421 pp., 2009.
- Borgman, C.: The Conundrum of Sharing Research Data, *J. Am. Soc. Inf. Sci. Tec.*, 63, 1059–1078, doi:10.1002/asi.22634, 2011.
- Burton, I.: Forensic Disaster Investigations in Depth: A New Case Study Model, *Environment Magazine*, 52, 36–41, 2010.
- Centre for Evidence-Based Conservation: Guidelines for Systematic Review in Environmental Management, Environmental Evidence, available at: [www.environmentalevidence.org](http://www.environmentalevidence.org) (last access: 21 June 2012), 2010.
- Cooper, H. M., Hedges, L. V., and Valentine, J. C. (Eds.): The handbook of research synthesis and meta-analysis, 2nd Edn., Russell Sage Foundation, New York, 615 pp., 2009.
- European Commission: Towards better access to scientific information: Boosting the benefits of public investments in research, European Commission, Brussels, available at: [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/era-communication-towards-better-access-to-scientific-information-en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information-en.pdf), 2012.
- Evans, J. and Reimer, J.: Open Access and Global Participation in Science, *Science*, 323, 1025–1025, doi:10.1126/science.1154562, 2009.
- Farace, D. and Schöpfel, J.: Introduction Grey Literature, in: *Grey Literature in Library and Information Studies* edited by: Farace, D., and Schöpfel, J., De Gruyter/Saur, Berlin/New York, 2010.
- Genova, F.: The many faces of SIMBAD, in: 16th Annual Conference on Astronomical Data Analysis Software and Systems, 15–18 October 2006, edited by: Shaw, R. H. and Bell, F., DJ, *Astronomical Data Analysis Software and Systems XVI*, Astronomical Society of the Pacific Conference Series, Tucson, AZ, 145–152, 2007.
- GO Consortium: The Geneontology, available at: <http://www.geneontology.org>, last access: 31 August 2012.
- Haendel, M. A., Vasilevsky, N. A., and Wirz, J. A.: Dealing with Data: A Case Study on Information and Data Management Literacy, *PLoS Biol*, 10, e1001339, doi:10.1371/journal.pbio.1001339, 2012.
- Haeussler, M., Gerner, M., and Bergman, C. M.: Annotating genes and genomes with DNA sequences extracted from biomedical articles, *Bioinformatics*, 27, 980–986, doi:10.1093/bioinformatics/btr043, 2011.
- Hannah, D., Demuth, S., Looser, U., and Prudhomme, C.: Large-scale river flow archives: importance, current status and future

- needs, *Hydrol. Process.*, 25, 1191–1200, doi:10.1002/hyp.7794, 2011.
- Higgins, J. P. T. and Green, S.: *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0 (updated March 2011), The Cochrane Collaboration, available at: <http://www.cochrane-handbook.org/> (last access: 21 June 2012), 2011.
- Houghton, J.: *Costs and Benefits of Data Provision*. Report to the Australian National Data Service, Centre for Strategic Economic Studies Victoria University, Melbourne, Australia, available at: <http://ands.org.au/resource/houghton-cost-benefit-study.pdf> (last access: 11 September 2012), 2011.
- Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D. P., Kania, R., Schaeffer, M., St Pierre, S., Twigger, S., White, O., and Yon Rhee, S.: Big data: The future of biocuration, *Nature*, 455, 47–50, doi:10.1038/455047a, 2008.
- Hübl, J., Kienholz, H., and Loipersberger, A. (Eds.): *DOMODIS – Documentation of mountain disasters. State of discussion in the european mountain areas*, *Interpraevent, Schriftenreihe 1, Handbuch 1*, Klagenfurt, 40 pp., 2002.
- IRDR: *Forensic Investigations of Disasters. The FORIN Project, Integrated Research on Disaster Risk (IRDR)*, Beijing, available at: [http://www.irdrinternational.org/wp-content/uploads/2012/2006/FORIN-REPORT\\_web.pdf](http://www.irdrinternational.org/wp-content/uploads/2012/2006/FORIN-REPORT_web.pdf) (last access: 11 September 2012), 2011.
- Luzi, D.: Grey documents in open archives, in: *Grey Literature in Library and Information Studies* edited by: Farace, D. and Schöpfel, J., De Gruyter/Saur, Berlin/New York, 2010.
- MacDonald, B. H., Wells, P. G., Cordes, R., Hutton, G. R. G., Cosarini, D. M., and Soomai, S.: The use and influence of information produced as Grey Literature by international, intergovernmental marine organizations: Overview and current research, in: *Grey Literature in Library and Information Studies* edited by: Farace, D. and Schöpfel, J., De Gruyter/Saur, Berlin/New York, 2010.
- Mackenzie Owen, J. S.: The expanding horizon of Grey Literature, 3rd International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, 13–14 November 1997, Luxemburg, 1997.
- Marx, V.: My data are your data, *Nat. Biotechnol.*, 30, 509–511, doi:10.1038/nbt.2243, 2012.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007WR006744, 2008a.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 2. Combining data evidence, *Water Resour. Res.*, 44, W08433, doi:10.1029/2007WR006745, 2008b.
- Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., and Huncheda, Y.: HESS Opinions “More efforts and scientific rigour are needed to attribute trends in flood time series”, *Hydrol. Earth Syst. Sci.*, 16, 1379–1387, doi:10.5194/hess-16-1379-2012, 2012.
- OBO Foundry: *The Open Biological and Biomedical Ontologies*, available at: <http://www.obofoundry.org/>, last access: 31 August 2012.
- Petrow, T., Thielen, A. H., Kreibich, H., Merz, B., and Bahlburg, C. H.: Improvements on flood alleviation in Germany: Lessons learned from the Elbe flood in August 2002, *Environ. Manage.*, 38, 717–732, 2006.
- Ranger, S. L.: Grey Literature in Special Libraries: Access and Use, in: *GL-Conference series, Sixth International Conference on Grey Literature: Work on Grey in Progress*, New York Academy of Medicine (USA), 182 pp., 2004.
- Renear, A. H. and Palmer, C. L.: *Strategic Reading, Ontologies, and the Future of Scientific Publishing*, *Science*, 325, 828–832, doi:10.1126/science.1157784, 2009.
- Rothstein, H. R. and Hopewell, S.: Grey Literature, in: *The handbook of research synthesis and meta-analysis*, edited by: Cooper, H. M., Hedges, L. V., and Valentine, J. C., Russell Sage Foundation, New York, 2009.
- Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bougueleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., de Daruvar, A., de Matos, P., Dix, I., Edmunds, S., Evelo, C. T., Forster, M. J., Gaudet, P., Gilbert, J., Goble, C., Griffin, J. L., Jacob, D., Kleinjans, J., Harland, L., Haug, K., Hermjakob, H., Sui, S. J. H., Laederach, A., Liang, S., Marshall, S., McGrath, A., Merrill, E., Reilly, D., Roux, M., Shamu, C. E., Shang, C. A., Steinbeck, C., Trefethen, A., Williams-Jones, B., Wolstencroft, K., Xenarios, I., and Hide, W.: Toward interoperable bioscience data, *Nat. Genet.*, 44, 121–126, 2012.
- Smith, M.: Communicating with data: new roles for scientists, publishers and librarians, *Learn. Publ.*, 24, 203–205, doi:10.1087/20110308, 2011.
- The Scientist: *Whither Science Publishing?*, *The Scientist*, 1 August 2012, available at: <http://the-scientist.com/2012/2008/2001/whither-science-publishing/> (last access: 10 September 2012), 2012.
- Thielen, A., Kreibich, H., Müller, M., and Merz, B.: Coping with floods: preparedness, response and recovery of flood-affected residents in Germany in 2002, *Hydrolog. Sci. J.*, 52, 1016–1037, 2007.
- Uhlemann, S.: Supplement to: Data Expansion: The potential of grey literature for understanding floods. Deutsches Geoforschungszentrum GFZ, doi:10.5880/GFZ.5.4.2012.001, 2012.
- Uhlemann, S., Thielen, A. H., and Merz, B.: A consistent set of trans-basin floods in Germany between 1952–2002, *Hydrol. Earth Syst. Sci.*, 14, 1277–1295, doi:10.5194/hess-14-1277-2010, 2010.
- Uhlemann, S., Thielen, A. H., and Merz, B.: A quality assessment framework for natural hazard event documentations: application to trans-basin flood reports in Germany, *Nat. Hazards Earth Syst. Sci. Discuss.*, 1, 143–192, doi:10.5194/nhessd-1-143-2013, 2013.
- Van Noorden, R.: Trouble at the text mine, *Nature*, 483, 134–135, doi:10.1038/483134a, 2012.
- Viglione, A., Borgia, M., Balabanis, P., and Blöschl, G.: Barriers to the exchange of hydrometeorological data in Europe Results from a survey and implications for data policy, *J. Hydrol.*, 394, 63–77, 2010.
- Warnick, W.: Tailoring access to the source: preprints, grey literature and journal articles, *Nature Webdebates*, available at: <http://www.nature.com/nature/debates/e-access/Articles/warnick.html> (last access: 12 July 2012), 2001.
- Weintraub, I.: The role of grey literature in the sciences, available at: <http://chaptercouncil.mlanet.org/roundtables/2006/GreyLit.pdf>, (last access: 9 November 2010), 2000.

- White, G. F., Kates, R. W., and Burton, I.: Knowing better and losing even more: the use of knowledge in hazard management, *Environmental Hazards*, 3, 81–92, 2001.
- Whitlock, M., McPeck, M., Rausher, M., Rieseberg, L., and Moore, A.: Data Archiving, *Am. Nat.*, 175, 145–146, doi:10.1086/650340, 2010.
- Wissenschaftsrat: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020, available at: <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>, Berlin, 90 pp., 2012.
- Wood, J., Andersson, T., Bachem, A., Best, C., Genova, F., Lopez, D. R., Los, W., Marinucci, M., Romary, L., Van de Sompel, H., Vigen, J., and Wittenburg, P.: Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data – A submission to the European Commission, European Union, available at: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf> (last access: 11 September 2012), 2010.