Article

Alpine Glaciology: An Historical Collaboration between Volunteers and Scientists and the Challenge Presented by an Integrated Approach

Laura Criscuolo [1,2,*]**, Monica Pepe** [1]**, Roberto Seppi** [2]**, Gloria Bordogna** [3]**, Paola Carrara** [1] **and Francesco Zucca** [2]

[1] Institute for Remote Sensing of the Environment, National Research Council of Italy, Via Bassini 15, I-20133 Milano, Italy; E-Mails: pepe.m@irea.cnr.it (M.P.); carrara.p@irea.cnr.it (P.C.)

[2] Department of Earth and Environmental Sciences, University of Pavia, Via Ferrata 1, I-27100 Pavia, Italy; E-Mails: roberto.seppi@unipv.it (R.S.); francesco.zucca@unipv.it (F.Z.)

[3] Institute for the Dynamics of Environmental Processes, National Research Council of Italy, Via Pasubio, I-24044 Dalmine (BG), Italy; E-Mail: gloria.bordogna@idpa.cnr.it

* Author to whom correspondence should be addressed; E-Mail: criscuolo.l@irea.cnr.it; Tel.: +39-2-2369-9456; Fax: +39-2-2369-9300.

Abstract: European Alpine glaciology has a long tradition of studies and activities, in which researchers have often relied on the field work of some specialized volunteer operators. Despite the remarkable results of this cooperation, some problems in field data harmonization and in covering the whole range of monitored glaciers are still present. Moreover, dynamics of reduction, fragmentation and decline, which in recent decades characterize Alpine glaciers, make more urgent the need to improve spatial and temporal monitoring, still maintaining adequate quality standards. Scientific field monitoring activities on Alpine glaciers run parallel to a number of initiatives by individuals and amateur associations, keepers of alternative, experiential and para-scientific knowledge of the glacial environment. Problems of harmonization, coordination, recruitment and updating can be addressed with the help of a collaborative approach—citizen science-like—in which the scientific coordination guarantees information quality and web 2.0 tools operate as mediators between expert glaciologists and non-expert contributors. This paper gives an overview of glaciological information currently produced in the European Alpine region, representing it in an organized structure, functional to the

discussion. An empowering solution is then proposed, both methodological and technological, for the integration of multisource data. Its characteristics, potentials and problems are discussed.

Keywords: glaciology; data quality; citizen science; volunteered geographic information; incidental information; data harmonization; data quality

1. Introduction

Glaciers and ice caps (excluding Greenland and Antarctica ice sheets and surrounding glaciers) cover on Earth an area ranging from 510,000 to 514,000 km^2 (smaller and larger estimate) [1]. The European Alpine glaciers represent about 0.5% of the glaciated land surfaces [2]. Even if limited in extension, they have always been an important reference for the world glaciology, being both the first to be studied and monitored continuously, and good indirect indicators (proxies) of climate change. Alpine glaciers are currently the subject of the longest and most reliable measurement series in the world and they are the focus of several experiments and projects of international relevance [3,4]. As is well acknowledged, the last century, and particularly the last 30 years, has seen a general trend of reduction in glacier surfaces and volumes, caused mainly by the increase in global mean temperature and a reduction in winter snowfall [5]. This has resulted in a dramatic regression of the Alpine glaciers, in some cases leading to the fragmentation, morphological mutation and full extinction of several glacial bodies [6–8].

Some international scientific organizations and initiatives—such as the World Glacier Monitoring Service (WGMS) and the Global Land Ice Measurements from Space (GLIMS)—assume leadership roles in the glaciological sector, and encourage communities to produce information according to common standards and frameworks. Nevertheless, national and regional authorities traditionally make their policies in the field independently from international initiatives, by their own means and resources [9]. In many cases, the initiative is committed by political and administrative authorities to voluntary associations and research groups, institutes and universities, which make use of their expertise at a national or local level, either collaboratively or independently. Generally, these organizations take charge of periodically providing measurements of physical and morphometrical parameters, sometimes complemented by remote sensing images and related interpretations. As a consequence, there is a great heterogeneity in the types of data observed, in their collection, processing, storage and dissemination procedures, not only from country to country, but often also from region to region within the same country.

Despite the importance of the work carried out in the European Alps in the field of glaciology, several factors make it difficult to build a comprehensive knowledge. Among them there are the location of glaciers in remote areas, the dynamic nature of the formations, the restricted period for observations (typically late summer), the scarcity and non-continuity of funds allocated to research and monitoring, and the variety of procedures for data collection and processing. All these factors contribute to the inadequacy in monitoring of the complex dynamics affecting the European Alpine glaciers.

In addition to these monitoring activities, conducted by experts in glaciology, remote sensing and geomorphology, there is a number of amateur observation activities, conducted by individuals or by groups of enthusiasts organized in associations, and whose background and expertise are extremely varied. They produce a significant amount of information, often published and shared through the channels of the web. Since this information is typically varied in type, characteristics and sources, it is usually neglected by the scientific world, due to its variability in format and content, difficult traceability, lack of information about accuracy, complex reliability evaluation, and complexity of privacy and copyright policies. On the other hand, such an extensive set of documents is potentially a valuable source of knowledge.

In this paper, we present and describe an approach aiming at organizing and aggregating both the traditional glaciological data and those observations, provided by alternative sources, usually ruled out of the scientific paradigms, and consequently scattered and unused. Systematically combining expert and non-expert glaciological information provides the opportunity to increase—both in time and space—the knowledge of Alpine glaciers, in terms of their current state and dynamics. The method by which this goal is achieved is challenging, and consists in a systematic and partially automatic aggregation of data from specialist, non-expert, and even unwitting sources.

To this end, the next chapter deals with the identification and analysis of available forms of glaciological data relating to the European Alps. They are characterized and represented on the basis of quantitative and qualitative attributes. In the third chapter, we identify the methodological and technical steps required to realize a system for the integration of multisource glaciological data. The most critical issues of the approach are discussed in chapter four, highlighting the possible solutions and providing recommendations. Conclusions are then drawn and perspectives are provided in the last section.

2. Glaciological Data: State of the Art

In scientific research, it is an accepted practice to rely on data derived from experts' activities and from sources conventionally believed to be authoritative. Alpine glaciological data for scientific issues are therefore selected according to criteria of authority, accuracy and adherence to procedures. An exception is represented by a particular category of field observations, produced by "non-expert" but "experienced" operators. In some areas of the Alps, indeed, periodic measurements of some glaciological parameters are usually coordinated by a scientific staff member, but handled by volunteers. The volunteer operators of the Italian Glaciological Committee are an example of this practice. Part of the annual measurement campaign on glacier length variations is committed to by them. These operators, often without a scientific glaciology background, undergo preparatory training and eventually acquire some competence. This enables them to produce observations with a certain reliability and consistency, and makes the information they provide useful material for scientific glaciology.

In addition, there are a number of data and information, often coming from outside the scientific realm, which has extremely heterogeneous contents and formats. The traditional glaciology usually does not examine, for example, the information produced by mountain amateurs, climbers, guides, nor by specialists in fields other than glaciology (photography, biology, meteorology, *etc.*).

In this context, a couple of elements acquire a key relevance: (i) the recent change in approaching geographic information, which has become more familiar to non-specialist users, and makes them aware of their dual role of user and producer (*produsers* for [10]); and (ii) the simultaneous development of collaborative and social technologies, which prove to be useful in collecting and managing informal geographic content from various sources [11,12], also within a scientific framework.

In recent years, some experiments have sprung up using geospatial web and non-expert contributions to collect information about the cryosphere. Extra-Alpine examples are the *Community Collaborative Rain, Hail & Snow Network* [13], originated at the Colorado State University to perform collaborative mapping of precipitations in the US and Canada, and the *Alpine of the Americas Project* [14], a public call to repeat historical photos of glaciers in America. Other initiatives are then referred to the European Alpine Area, such as those coming from some Italian Regional Agencies for Environmental Protection, which provide citizens with geoportals and mobile applications for the participative collection of local snow-cover information [15,16], and the "Ghiacciai di una volta" project, promoted by the Science Museum of Trento, for repeating historical photos on Italian Glaciers [17].

In all these initiatives, however, the collaborative aspect of the collection comes in specific forms of participation—some of them taking into account authenticated expert measures only, some others volunteered observations only, and other ones being mainly informal contributions in the form, for instance, of RSS feeds. No case has been found in which a real integration has been performed among different sources and forms of glaciological, snow and ice data.

As previously introduced, data related to Alpine glaciers, potentially useful for glaciological research, are numerous and extremely varied. They include observations and measurements, descriptions in free text, codes, conventional signs, maps and graphic works. These data are variously combined, produced for several purposes and in different contexts, and have highly variable sizes and readability. In this work, we attempt to organize these diverse realms of data in large homogeneous groups, which allow a more effective treatment. In order to derive useful information for the data categorization, first of all, we recall some concepts and explain some terms, which recently came out of ordinary usage.

Data commonly considered "official" are a set of measures, observations and elaborations from an authoritative source, scientific or specialist, obtained within a well-framed working environment (a research project or a business) through acquiring and processing procedures that are well-defined and documented. Since the last decade, even the wide-ranging set of "unofficial" data, coming from the activities of non-authoritative contributors, started to arouse some interest inside the scientific community. This set of activities and information is defined with different terms, such as *Neogeography* [11,18], *Citizen Science* [19,20] and *Volunteered Geographic Information* (VGI) [21]. These labels, and especially *VGI*, well fit the contributions coming from the observation campaigns of glaciological operators, which have a clear scientific purpose, a voluntary nature and geographical-related content. The term "volunteered" refers to actions performed in a conscious, deliberate way, without a personal—typically economical—reward. However, it is improper to assert that all publicly available non-expert data, potentially of interest for science, are "volunteered". In fact, the act of providing public information, related in this case to glaciology, is neither always aware nor free of remuneration. In publishing any of this information, the author is in fact conscious of providing her/his observations on public web pages, but she/he is often unaware of the contribution in terms of

geographic information, and glaciology, that it contains. Even the unrewarded nature of these contributions is not always obvious, as the action of publishing some information can bring to their providers some benefits—economical or professional—that are conferred to them independently from the contribution unconsciously given to science. These non-expert and non-volunteered contributions need, therefore, another denomination. We suggest the more appropriate name of "incidental data", to which corresponds the analogue concept of "incidental information". The term "incidental" refers here both to the fortuity of the scientific relevance of the contribution, and to its minor worth for research.

3. Materials and Methods

3.1. Definition of Available Data Resources

Before describing an integrated system for managing heterogeneous glaciological data, we need to categorize the data resources playing a role in it. In order to do this, we firstly focus on three influential criteria:

- A—Expertise: the specialization of the author or of the context in which data are created;
- B—Intentionality: the awareness of the relevance of data for science in general or for a research project, and therefore the willingness to cooperate;
- C—Reward: the benefit assigned to the author for the data distribution.

We can now define three data categories, obtained by combining these criteria:

- Official data: expert, intentional and rewarded
- Volunteered data: intentional and unrewarded
- Incidental data: (case 1) unintentional
   (case 2) non expert, intentional and rewarded

Most of incidental data for glaciological research fall within case 1—unintentional. A small number of data, very uncommon in the Alpine realm, fall within case 2. These data are not official, being non expert, and are not volunteered either, being rewarded. They can be considered incidental by extension, by reason of their minor worth for research.

These categories can be effectively represented according to Set Theory (Figure 1), as they are unambiguous and free from overlapping. They are useful in describing data resources even beyond the glaciological realm. This theoretical framework will be the premise for the following discussion.

3.1.1. Official Glaciological Data

Official glaciological data, as already discussed, are acquired or processed by experts under well-documented procedures, often standardized. Typical producers are research institutes, Universities, local authorities, cartographers, business companies owning detection equipments or managing remote sensing images. The benefits assigned to data authors for their publication may be a direct compensation for the service, a regular salary, or a non-cash prize such as a score or a positive evaluation that brings to author professional advantages.
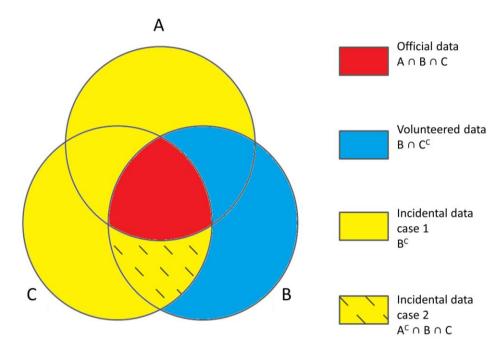
Figure 1. Representation of official, volunteered and incidental data according to Set Theory.

Data and products acquire in this way an economical value, which makes them sometimes distributed with a fee or under trade agreements. In some cases, on the contrary, these data are freely provided, usually via the web. The choice is usually a consequence of opening data policies (e.g., Open Data movement for public administrations) and is also supported by the European Directive INSPIRE [22] and its implementation of national decrees. These initiatives are pushing governments and private entities to share via the web their geographical information in standard forms.

The web distribution of official data often takes place through publications in scientific journals or technical reports. Data and their geographical elaborations can be also disseminated through institutional geoportals, in the form of web services or GIS layers.

Field data most commonly collected in alpine glaciology are measurements of the glacier length variations and of the snow cover thickness. Both measurements can be performed with different methods, depending on the working environment, the team's equipment, and the aims of the research. The first measurement, for example, may range from repeated tape readings of the distance between the glacier margin and landmarks on the glacier forefield to the use of devices such as total stations or GPS. The second one can be executed by reading the height on a pole plunged into the snow cover, or even by performing complex surveys with radar, laser or other geophysical techniques.

Other important data come from remote sensing, including satellite imagery, LIDAR and photogrammetric air surveys. By integrating field and remote observations, glaciologists assess area and volume variations and surface mass balance. Those data are often used as inputs for hydrological modeling and forecasts on the evolution of the freshwater reservoirs represented by glaciers.

These data, because of their nature and purpose, present recurrent forms and content, often organized in homogeneous structures. This means that they can be compared each other, consistently placed in series and collections, and interpreted without misunderstanding.

The authority of these data is guaranteed by the professional background of the sources, and their truthfulness is subject to review by the scientific community. In this context, in fact, the reliability of

the information is supported by the reviewing process, as common practice in scientific publications. This working method strongly motivates authors to maintain high quality standards, a necessary condition to get credibility and authority [23].

3.1.2. Volunteered Glaciological Data

Volunteers performing tasks for Alpine glaciological monitoring come from very different backgrounds. Sometimes they are extemporary collaborators, but, more often, they are persons who spontaneously take charge of observing glacier changes after having become acquainted with monitoring initiatives and having appreciated their aims. In such cases, volunteers are usually members of a group and are guided by a scientific coordination. The coordinator draws the guidelines for the group, rules the content creation, and sometimes takes care of volunteers' training by means of manuals, booklets, and specific courses.

The scientific supervision can be more or less pronounced during the data creation; anyway it significantly influences the production of high quality datasets. In the reporting phase, volunteers are sometimes supported by assistance tools for the compilation (charts, codes, protocols), and by revision mechanisms to amend the collected data sets. The approach to quality checking of volunteered data is a choice of major concern for the final scientific use.

Data are usually distributed using web spaces and applications, in which the collaborators (glaciologists or operators) can sometimes fill out web forms, or more often upload pictures and documents, exchange hints, and share content and experiences.

Volunteered glaciological data, as explained before, are frequently exploited by research projects or addressed to monitoring initiatives of public interest. In such cases, the publication of volunteered data takes place in line with the data policy of the project, as established by the institutional/scientific coordination.

Volunteered glaciological data are often complex hybrid objects, combining images (mainly pictures), textual annotations, numerical observations and, sometimes, geographical features (e.g., GPS points in GIS formats).

Typically, volunteers provide pictures of Alpine glaciers, often shot from established field landmarks, following a pre-set direction (azimuth) and repeated in time (on a yearly basis), in order to create a series of visual observations of a glacier taken in the same conditions. In the last few years, pictures are taken by digital cameras and may include some metadata useful for describing the observed scene (GPS coordinates, date and time). In most cases, pictures have documents annexed, which report details such as author, name of the glacier, name of the location or landmark code, azimuth, focal, meteorological conditions, and notes by the author.

Free textual contributions from volunteers can report observations about the state of the glacier as well as particular phenomena, or detail pictures and measurements. They can present very different forms, syntax and contents.

The numerical contributions can include measurements of glacier length, snow thickness, position and elevation estimates for significant glacial elements.

Contributions in the form of spatial features are less common. The reasons for that can be found in the lack of technical equipment and, often, in the limited know-how of the volunteers, who cannot be

expected to own adequate tools and methods for cartography. The spatial features are typically vectors (stored as SHP or KML files) laying out points and borders meaningful in the glaciers description.

Recent developments in smart technologies could predict future scenarios in which light mobile devices (smartphones, tablets) will substitute, by means of apps and devices, the traditional paper annotation for recording glaciological observations in the field, and a valuable aid to integrate them in on-line processes and archives. Smart devices are already in use in the field of geography as well as geological surveys, and, in some advanced frameworks, are used to collect different kinds of observations and to connect to remote databases and on-line applications. They can provide user-friendly interfaces and interactive tools, helping non expert cartographers to collect and map geographical information.

3.1.3. Incidental Glaciological Data

Following the volunteers training, operating and compiling phases is quite demanding and thus, the number of volunteers contributing to glaciological research is still small as compared to the whole crowd of amateurs used to share glaciers' data on the web. Despite their limitations, incidental data cooperate in the understanding of phenomena and significantly improve the frequency of observations. This fact makes incidental glaciological data an interesting informative complement for alpine glaciology.

Providers of contents that are incidentally of glaciological interest do not belong to a single category. For example, incidental information can be provided by a civic employee, a climber, a mountain hut keeper, or a student. Their observations can sometimes unintentionally help the data validation, open up new visions about phenomena, or collaborate to monitor trend indicators. This huge variability of sources and approaches makes it particularly difficult to manage and assess this kind of data.

Incidental glaciological data present hybrid forms even more often than volunteered information: images, videos, annotations, measurements, spatial features are usually combined in heterogeneous and variable structures.
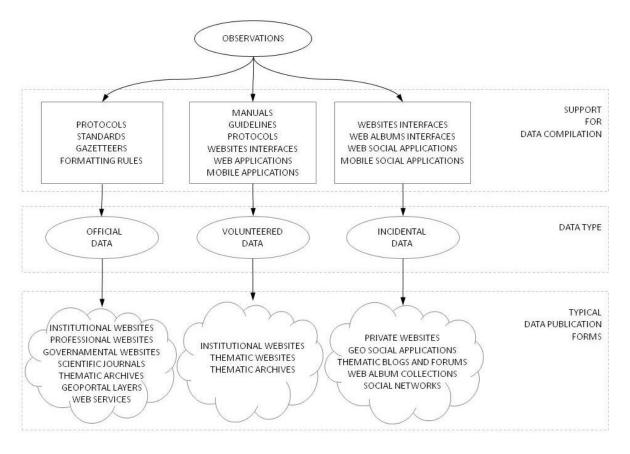
Contributions are often produced in the form of trip-reports, published on the web by hikers and mountaineers. They frequently provide information on snow and ice conditions, and the presence of particular formations or risks (hanging blocks, clefts, crevasses, *etc.*). Photo collections are other typical incidental products potentially useful for glaciological research. These photos are often shot to document expeditions or trips and, when shared on the web, can provide information on geological, biological and hydrological conditions. Other ancillary and incidental information for glaciology can be produced by local authorities and web magazines, spreading on the web local news, among which rock falls, accidents, avalanches, and extreme meteorological events are eventually of interest for glaciologists.

All these types of incidental information are usually distributed via the web by means of forums, blogs, web pages, web photo albums, and social and geographic applications (for instance, applications for social mapping or virtual globes). They are commonly accessible to the general public for free, and are frequently provided with social tools for sharing, commenting and rating content. Incidental data are often accompanied by scarce or ambiguous metadata. Sometimes they even accumulate uncertainty

during phases of editing, social sharing and commenting. This makes it difficult to retrieve and interpret information such as authorship, geographical position, date and time, and distribution rights.

The characterization of different workflows, as described in the previous paragraphs, is synthesized in Figure 2.

Figure 2. Synthetic representation of workflows for official, volunteered and incidental data.

3.2. Managing Glaciological Heterogeneous Data Resources—User Requirements Overview for Integrated System

In Alpine glaciology, managing heterogeneous information distinct with respect to semantics, nature, format, and sources characteristics is desirable. In fact, there are two main strong reasons to adopt a solution capable to manage such information. A first factor is the interest, expressed by both institutional and private subjects, to monitor a large and remote landscape. Such interests are not supported by adequate funding for monitoring. This makes it necessary to optimize all available resources by stimulating collaborations and involving as many volunteers as possible, scattered along the Alpine valleys. The second factor is that if, on one side, there is a strong interest by a lot of people caring for mountains, on the other side, there is not a channel for communicating glaciological information to the general public in a simple, clear and easily-accessible way; there is a lack of a common "space" where all interests could find their place. Moreover, professionals and stakeholders could share the need for an integrated system to both access and analyse this kind of information too, since there is not yet a gateway to obtain up-to-date information about the state of glaciers in the whole Alpine Chain.

3.3. Architecture Overview of an Integrated Data Management System

A system for multisource glaciological heterogeneous data management has to be designed so as to address all the described needs properly. Basic functionalities of the system, as conceived in the field of glaciology, shall include: collecting thematic material; supporting data storage and management in homogeneous and acknowledged formats; enabling the dissemination, in a clear and handy way, to a general, heterogeneous and wide public; and promoting data processing and analyses for professionals and stakeholders.

We propose an effective multisource, collaborative approach for alpine glaciology, described here by its technical steps. Each step, from the data retrieval until the final exposure, is graphically described in Figure 3a, together with the components of the system (in Figure 3b).

The earliest data input phase is carried out off-line by the system and asynchronously with respect to the subsequent discovery and access phase, and can be reiterated with a given frequency depending on the rate of updating and creating the information through the monitored sources.

The data input is performed by executing four subsequent processes: the crawling of data published on the web, the metadata creation, the data validation on quality bases and the final organization of data into a database.

The crawling consists of visiting a portion of the web by starting with well-known and authoritative sources' repositories, besides interesting repositories potentially rich of glaciological information (*i.e.*, from known URLs) to fetch the web pages in order to extract or create from their contents the metadata (A in Figure 3).

The crawling of authoritative source repositories, such as those containing official data, can be done with the use of a simple crawler, since we know that information there is structured by a known schema specific for the dataset, archive or literature repository crawled. Volunteered data are often organized into semi-structured repositories, made available by research project managers. They can be easily visited and fetched by a simple crawler. In other cases, volunteered data are not coordinated and thus, are scattered in non-authoritative web sources, together with incidental data. For such non-authoritative sources, we need a focus a crawler that filters only the subsets of the web pages' contents that can be of glaciological interest.
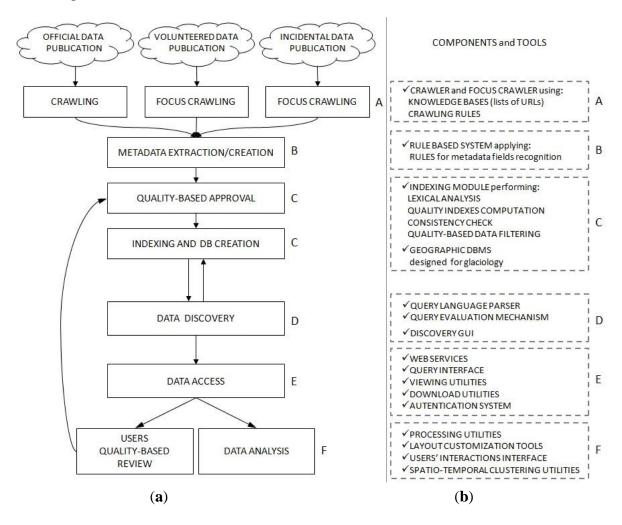
The crawler visiting official and volunteered data, whose repositories and (semi-) structured schemas are known, can apply rules that select only relevant fields within the structured data, such as measurements, observations, images and graphs. The focus crawler, that must identify relevant non-structured data (mainly incidental and occasionally volunteered), must be defined with a set of heuristic rules that select web pages containing specific terms, such as names of Alpine glaciers in the caption of images, and glaciological domain specific terms, drawn from an ontology.

Once the web pages containing the information of interest are identified and selected, their contents need to be analysed to extract or create the metadata [24]. This creation of metadata can also exploit semantic annotations, tags and appreciation ratings associated with the contents of the web pages [24] (B in Figure 3).

To this end, techniques of lexical analysis, natural language processing and text mining can be useful. The exploitation of multiple techniques has the aim to recover the largest possible number of metadata, extracted from explicit and structured data (bounding box, authorship, date and time,

lineage, *etc.*) and from implicit ones (toponyms, tags, users profiles, links, addresses, *etc.*), since this has an impact on the final quality of the information [25]. The extraction of spatial and temporal metadata can be done automatically or semi-automatically from images [26], free text, keywords and tags [27,28], or textual structures, encoded for social applications, such as Twitter tweets' components [29].

Figure 3. Representation of the proposed methodology: (**a**) functionalities and (**b**) components.

(**a**)   (**b**)

Once the metadata are generated their quality must be checked for approval. As discussed later in this paper, this is a delicate matter, specifically when considering volunteered and incidental information.

The techniques applied strongly depend on the type of source. The quality of official information is usually assessed before publication and the pertinence to the glaciological research needs to be estimated. On the contrary, we need to evaluate validity of volunteered and incidental information, together with the reliability of non-expert contributors (C in Figure 3). The quality evaluation can be carried out by computing appropriate quality indexes for each type of metadata field (authorship, timestamp, geographic footprint, *etc.*) following criteria such as completeness, correctness, accuracy, intelligibility and consistency. Only the volunteered contributions whose quality indexes exceed fixed minimum thresholds will be approved and retained [30]. Finally, to assess the quality of the collected incidental information in most cases, we will need to complement the methods used for volunteered information with the aid of a human moderator, who manually assigns quality scores to the

contributions. Incidental contributions whose quality exceeds the minimum threshold will be approved and retained.

The cleansed metadata can finally be indexed and organised into a geographic database so as to be available for the discovery phase (C in Figure 3).

All the above processes should be periodically re-executed, with a frequency that should be determined as a function of the life cycle of information (creation, revision and deletion date). Sites, which are updated very often, will need to be frequently visited by the crawler.

The database should always track the data source and the author of every single piece of information to allow for the determination or estimation of reliability at any time. Meanwhile, it is important to properly represent the data sources in order to estimate their authoritativeness, popularity and influence. Such ancillary data could be of use to improve the quality validation performed in subsequent reiterations, by removing inaccurate or intentionally incorrect information, as it is often generated by scarcely reliable sources.

The indexed information shall be provided in different ways in order to enable and facilitate its discovery and access. To this purpose, different discovery approaches can be adopted, based on filtering, retrieval, or browsing techniques. In a filtering (or push) approach, a selection of information is periodically fetched to the user's own address, according to his/her preferences in regards to content and frequency. In retrieval (or pull) approach, the system interprets explicit user's queries and retrieves the corresponding information items. In the browsing approach, a client browser assists the user navigating through clusters or classes of information items in case it is organized into hierarchical trees. All these alternative techniques need a query language parser to interpret the users' preferences, and one or more graphic interfaces (GUIs) to assist users in the queries composition, in the content navigation and selection (D in Figure 3). Queries and preferences can be expressed in several forms: natural language texts, controlled terms from a dictionary of indexes or from a thematic thesaurus, toponyms selected from a geographic gazetteer, spatial coordinates or bounding boxes, time spans, or even complex queries containing Boolean and relational operators, expressed in a formal language (such as SQL, Xquery, Xpath, SPARQL, *etc.*).

Discovered information must be presented to the user in an easy accessible way. For this to be achieved, they should set up appropriate web and mobile interfaces where users can examine information (E in Figure 3). The information items can be represented by means of icons and styles to make clear the categories of their sources—official, volunteered or incidental—and their overall quality scores. Glaciological information items usually contain one or more spatial references, expressed by coordinates or toponyms. This allows the display of this set of information items on a interactive map, where they can be represented through styled markers, placed on the centre of the geographical footprint. The map viewer will provide zooming and panning utilities, while a querying and filtering tool can allow textual and visual queries by clicking on the map. A geographic gazetteer coupled with (reverse) geocoding utilities can allow automatic translating of toponyms in the corresponding geographical locations. A further accessing mode, suitable for geographical and non-geographical items, can display the lists of retrieved information, sorted by temporal and/or quality criteria. Even in this presentation mode, the user shall be provided with querying and filtering utilities.

The retrieved items can be summarized or conflated by some of their characteristics, such as category of source, time of creation, quality indicator, and relevance to the query.

In order to protect private data and better manage the user's access, it could be necessary to activate an authentication system, and set up read and write rights.

Data hosted on the server of the project can be offered to users also allowing its download. Downloads shall comply with copyright and licensing policies, together with intellectual property rights (see discussion later in this paper).

The formats available for the download shall be as compliant as possible to the most diffused applications in technical and amateur contexts in order to achieve the best information exploitability by different end-users. Some simple additional functionality shall be implemented within the system to improve the exploitation of information; for example, GIS-like tools that overlay concurrent data layers, customize layers' styles (colours, symbols, transparency, *etc.*), select and zoom on particular themes (attributes), or geographic and time windows to perform simple spatial and geo-statistical analyses on selected data (F in Figure 3). Finally, some utilities can be implemented in order to promote a collaborative quality assessment on presented data. The users' community can be provided with social tools, by which comment and rank contributions, assessing them for consistency and reliability criteria, point out irregularities and discrepancies. Feedback obtained by this collaborative effort could enter the quality validation process by iterative learning.

4. Discussion about Critical Aspects

The integrated management of glaciological information generated by distinct data sources offers indisputable advantages, although it introduces some concerns, mainly in regards to the quality of data, spatial and language features, protection and copyrights, and web users' involvement.

4.1. Data Quality

Quality aspects that mainly affect glaciological data usability are several, among which accuracy and precision of geo-location and of observations, completeness and intelligibility of contents, as well as the reliability of information and the trustworthiness of the data source.

In the literature, there are useful references about information quality modelling and assessment, even in the case of Geographic Information (GI) and VGI [31].

The quality assessment is a step that cannot be avoided or underestimated; it is necessary to have a strategy regarding quality policy in order to, at first, properly manage the information, and then, to exploit it consistently by visualization, spatial analyses, modelling, *etc.*

Validity/usability of the information content and its credibility are the two fundamental criteria by which data quality can be appointed.

The information content validity, known also as intrinsic quality [30], depends on a combination of factors such as lineage, positional accuracy, attribute accuracy, logical consistency, and completeness [32], which, as a whole, make data fit for a given use [33]. It is therefore dependent on contents' inherent characteristics.

Methods applicable to audit these quality features could include ex-ante and ex-post techniques, combined in different ways. Ex-ante techniques act by preventing the creation of erroneous

contributions and guiding the contributor in providing effective data. Some examples of such techniques are the guided filling of protocols, the use of web-forms with fixed fields, the use of metadata automatically created by the measuring device (*i.e.*, GPS information associated to a picture taken with a smartphone), the use of ontologies and geographical gazetteers [34,35], the selection and training of volunteer contributors [36].

On the contrary, ex-post techniques are applied during fixing operations for the already created content, amending the defective components or sorting out inputted data by quality effectiveness criteria. Several examples of this kind of strategy are reported in the literature; for example, the huge databases of projects such as eBird [37] and FeederWatch [38] are processed by geo-statistical filters automatically, but also by human experts, in order to detect biases and maintain the data consistency [39].

In regards to the credibility of geo-information, on the whole it can be stated that it depends on both the trustworthiness and the expertise of the author [40] and that only a combination of these two aspects could assign credibility to the information [23,41].

The fundamental concept of credibility well suits both the conventional production of expert scientific information, and user generated content (UGC), even if the latter is complicated by some aspects specific to the web domain, like the difficult traceability of authors, the lack of standards and meritocratic selections, and the costly search for sources. In the last decade, several studies have been focussed on building credibility models [41], analysing quantitatively and qualitatively user generated content fluxes on the web by discussing their intrinsic characteristics, sources, subjects, drives [42–44], and the issue is still open and debated.

In the literature, we reported strategies and procedures aimed at the management of the quality of non-expert geo-information, as collected in specific project frameworks with the purpose to be used together with authoritative information. This is the case of Huang *et al.* [45], who propose a novel reputation system that makes use of the Gompertz function for computing device reputation score to estimate the trustworthiness of the contributed data.

Reddy *et al.* [46] developed specific metrics to quantify participant expertise and participation. De Longueville *et al.* [47], and next, Ostermann and Spinsanti [25] addressed the problem by proposing a workflow, integrated with existing Spatial Data Infrastructures, for automatically assessing the quality of VGI. Differently from Huang, these authors designed a complex procedure that, through several and iterative steps, assesses quality by considering not only trustworthiness but also relevance and completeness of the resource geographic content and of the related metadata.

It is commonly acknowledged that official data have a greater concentration of information reliable and useful to science, while volunteered and non-specialist data are more affected by inaccuracies and contain less scientific value.

Some authors have spent efforts to prove such a conjecture by analysing datasets of volunteered and specialized observations. Dickinson [39] reports a series of studies in which variations in observer quality are correlated to the author's preparation. Among factors influencing such variations are background and experience [36] together with the type of task [48–50], level of training, the company of a specialist in the field [51], and age and education of the author [52].

Despite the adverse qualitative characteristics of volunteered and incidental data, the larger number of potential contributions constitutes a significant strength for those data types.

Several studies have shown that the creative, aggregate use of non-expert contributions can generate new valuable information [47,53,54], and have documented situations in which local knowledge or expertise provide information of greater value than the expert knowledge [55]. There is evidence of the high potential of volunteered geographic information when collected and managed in well-structured contexts; also in the results of the analysis conducted by authors such as Haklay [56], Girres and Touya [57], Cipeluch *et al.* [58] who have evaluated the accuracy of OpenStreetMap data against reference sources.

The assignment of authority in the traditional expert-driven information is reached with an authoritarian, top-down, model. On the contrary, in non-expert, user generated content, the assessment of reliability follows a democratic paradigm according to a bottom-up model [59]. The combination of the two methods, however, is not only possible, but also can produce remarkable results. In this context, we should not underestimate the power of the web as a meeting place for participative evaluations: the continual access to the web content by a hybrid team of experts, locals, amateurs, and occasional visitors authorized to assess it, which may give rise to a sort of crowdsourced credibility assessment with a high potential for selection and judgment [23,60].

4.2. Spatial Domain and Language Policy

A second issue to be addressed in implementing a glaciological integrated system, is related to the choice of the spatial domain for data collection and retrieval, and, closely linked to it, the language policy of the entire system, including the natural language for querying.

Ideally, the optimum would be reached by extending the data retrieval to potential sources of glaciological information all over the world. In reality, actuating this ideal goal would become more of a problem. Firstly, it would require focus crawling instruments capable of interfacing with every language and thesauri to manage terms and concepts coming from the different geographical communities. Secondly, the consultation by users of such multilingual dataset strictly depends on the languages the user knows, and so it could be seriously hampered by the language choice. Even assuming the use of cross-lingual information retrieval, the level of intelligibility, and therefore of understanding, of the information would suffer an inevitable and not-quantifiable fall.

Thirdly, the retrieval of European Alps glaciological data from all sources of the world would entail a significant effort compared to the limited amount and pregnancy of recovered data.

The alternative consists in adopting one or more official languages (e.g., English, Italian, French), and limiting the sources for data crawling according to administrative, geographical or thematic criteria (e.g., Europe, Alpine Region, *etc.*).

This scenario involves other issues that need to be evaluated too. When adopting an official language one needs to consider several criteria such as ease of querying, maximization of fruition of information by potential users, maximization of ease for creation of high quality information. All of these criteria can determine distinct choices that can be conflicting with one another.

4.3. Data Protection, Copyright and Related Rights

The problem of protecting personal data arises significantly with the advent of web 2.0, the diffusion of mobile personal communication devices (MPCDs), and the increase in the number of

applications for sharing online multimedia materials collected by web users (web albums like Flickr, Panoramio; collaborative mapping platforms such as Google Earth and Google Maps; video sharing sites like YouTube; social networks like Facebook, MySpace, *etc.*). These technologies have also amplified the issues of managing privacy protection, leading, on one side, to the request for guidelines from the scientific and professionals' communities [61], and, on the other side, of claims and legal actions taken by the injured (a famous example is the number of protests and subsequent fixings introduced by the imaging activities of Google for its Street View application). Observations, either volunteered or incidental, particularly when collected by MPCDs, could contain private information, while they produce a tracking of user's locations and knowledge of his/her personal data, or data regarding places or subjects sensed, which gives rise to the need for data protection solutions [62].

The system we propose shall be able to deal with personal data protection issues in the three different given cases of official, volunteered and incidental data.

A solution for volunteered data directly inputted in the platform is the sign (also made implicitly) of an Informed Consent Form. The same solution neither applies for official and incidental data belonging to other sources of information, nor shelter from the risk to publish private or non-disclosure information (the "second hand smoke" effect in [63]).

To reduce the risk of publishing accidentally private or restricted contents, some authors applied algorithms to anonymize contributions [64], as well as to downgrade or filter out personal information [46,65]. However, such precautionary solutions have the side effect to discourage participation and disregard the intellectual properties of contributors.

One of the main obstacles when dealing with privacy protection in the case of incidental data is that, differently from the case of official data, metadata recording authorship and observation context are often lacking, completely missing, uncertain or in a format which is difficult to process (links to other web contents, nicknames, tags, *etc.*).

The same problems affect the retrieval of the copyright information regarding data. Such information is provided in very different formats by websites, forums, web galleries and any other web application giving access to data. Cascading links and cross references through web pages make it very difficult to trace original copyright information, which can easily be lost. The system shall address this problem by retrieving the constraints of data on originating web sites (thematic sites, catalogues, forums, web albums, *etc.*), then by respecting the data policies as determined at the web source.

Actually, the obstacles encountered in the retrieval and correct interpretation of constraints regarding data could be such an impediment to suggest a selection of data sources ruled by the distribution policies of web sites.

The management of Intellectual Property Rights (IPR) is another delicate subject in projects in which sharing of scientific data and knowledge is considered. Several experiences have proven that scientific data sharing on the web could be managed without debasing IPR and, rather, could enhance it, benefitting the authors or groups who contribute.

For example, in projects like EnvEurope [66], or within research institutes like NERC [67], policies for data sharing have been set up, which on one side, commit authors to publicly share their data, and, on the other side, require users to cite credits and offer possibilities for authors to join any related projects. In this way, the authors are encouraged to voluntarily share data on the web platform.

In the glaciological integrated system, volunteer contributors will be offered to choose among different data usage policies (or licenses) to be associated with their own contribution. There are some examples of this approach: in FLUXNET [68], authors can choose among three different Data Usage Policies: (i) access reserved to contributors (referred as LaThuile); (ii) access based on scientific proposals (Open Data); (iii) freely-distributed (Free Fair-Use).

It has to be reminded that IPR protection is a matter of law, regulated differently in each Country. The system is anyway supranational by definition, both according to the geographic entity—the Alps—that crosses different countries, and to the locations of the collected information of such geographic entities that are scattered on the World Wide Web.

The reference guidelines in this case could be the ones introduced by the World Intellectual Property Organization (WIPO), possibly refined by the European legislation on copyright and related rights (in particular by [69–71]).

4.4. Strategies for Participation

The system can act as a catalyst in promoting participation and fostering knowledge sharing. For this to happen, the system must be designed strategically according to three perspectives: appearance, quality, and quantity.

The appearance of the graphic user interfaces for both information creation by volunteers and visitors' discovery are crucial for its usage. It should communicate since the first sight scientific relevance and authority; this requirement should coexist with a user-friendly interface neither scaring off potential users, nor discouraging potential contributors. The page settings should facilitate visitors to find all information necessary for a deeper understanding of the content, avoiding at the same time to provide light users with unwanted information.

To make the user feel comfortable, well-known GUI, such as those of the most popular web mapping sites (Google maps, Google Earth, *etc.*), could be adopted. Even the data structure should result from a balance among criteria of completeness, readability and usability. The data entry procedure (the number of requirements in the data entry procedure) should be flexible to best adapt to skills, needs and purposes of contributors. Ad-hoc web forms can be used to provide support in the compiling phase, leading contributors to enter accurate and readable data, and suggesting to them standard terms from a shared lexicon. To this aim, checklists, multiple choice menus and optional fields can be used. Self-assessment tools can help providers by indicating the reliability degree of the entered contribution and declaring their own confidence level.

Moreover, some experiences and social science theories suggest that showing authors external perspectives on the value of their contribution can encourage participation [72].

With several solutions, it is then possible to emphasize the importance of voluntary contributions to the success of the project and ensure that the experience of contributors is satisfactory; for example, by communicating regularly to the volunteers' community the achievements of the project, or highlighting particularly significant contributions, or by publicly welcome the registration of new volunteers.

Other measures that help to motivate the volunteers are those which act to expand the user's possibilities of interaction with the web platform, for instance by sharing content with social networks,

or by using alternative communication channels such as e-mail, apps for smartphones or SMS (and geoSMS).

A reward system is another effective mechanism in supporting the participation of volunteers. Coleman *et al.* [43], analyzing the reasons which lead users to produce information on a voluntary basis, show that the "Social Reward" and "Personal Reputation" are among the main factors stimulating participation. Systems of social rating (thumbs up/down, star rating, *etc.*), voting, role upgrade based on user activity can be used to push these factors and encourage participation on the platform.

With regard to the participation of experts and specialists as volunteers, their collaboration can be encouraged—as already discussed in a previous paragraph—by a data policy which provides for intellectual property rights and encourages data users to cite authors or to involve them in their projects (this is, for instance, what happens in a project such as EnvEurope [66] or FLUXNET [73].

5. Conclusions

Glaciological data collected in the European Alpine domain are crucial for monitoring and understanding global climate change and related phenomena. In fact, even if European Alpine glaciers constitute a small amount of the whole glaciated land surfaces of the planet, due to their small size, they play the role of rapid response proxies and indicators of global changes.

Despite their importance, glaciological data sets are often small as compared to their environmental relevance; this is mostly due to the remoteness of areas to be inspected, and the lack of both sufficient financial resources and common strategies for glacier monitoring of the Alps as a whole. This is carried out by way of efforts and policies that are usually set up at a national or even local level, resulting in collections which are different in both observed parameters and surveying protocols. This scenario is further complicated by the fast dynamics of glacier bodies, which require a high frequency of surveys that is hard to perform. These reasons historically made voluntary associations, together with the scientific community, an important and active part of the observing and monitoring scene, with a large amount of glaciological information fully collected by volunteers, especially in Italy.

The fragmentation and heterogeneity of glaciological observations resulting from this scattered observing system is very high and requires strong efforts of harmonisation and pre-processing in order to achieve a comprehensive knowledge and analysis phenomena at a regional Alpine scale.

In this context, the recent outburst of a novel collaborative geospatial awareness, empowered by the web 2.0 technologies, introduces the opportunity of collecting and managing a lot of informal geographic content from various sources.

In this work, we proposed a system for multisource, heterogeneous information management to organize and aggregate both the traditional glaciological data sets and observations provided by alternative sources, out of the scientific realm, which are scattered on the web and currently unused.

To this aim, we have firstly identified and categorized the different data sources useful for glaciological research and monitoring on the basis of expertise, intentionality and reward. This led to the classification of glaciological data as official (expert, intentional and rewarded), volunteered (intentional but unrewarded) or incidental (either unintentional or non expert, intentional and rewarded).

After that, we defined a workflow in which methodological and technological solutions are proposed for the identification and management of glaciological, multisource, heterogeneous data.

At this stage of the research, some components of the described system have been designed and are under development. Among them there are a geographical-glaciological gazetteer, a knowledge base for focus crawling, and a quality indexing metric. A prototypal system has been set up for testing, viewing and querying use for volunteered and incidental information, restricted to three Alpine test glaciers. Future goals will regard the components' completion and the final assembly in a comprehensive system.

Incidental contributions, which have been mainly unexploited until now, could offer new information that otherwise remains hidden and unacknowledged, and could lead to new challenges in research by comparisons and analyses among such data, and traditional ones could introduce new application fields by the intersection of different disciplines: natural sciences, computer sciences and social sciences.

The use of an integrating system, as a favoured channel for data entry and access, could produce also the opportunity for the creation of a collaborative network for the reinforcement of existing communities and for the involvement of new subjects with the roles of volunteers, coordinators or providers.

Other benefits could be gained from the dissemination and diffusion of scientific information, which could be spread by original and customized ways, in order to reach, at different communication levels, both the community of experts in the field, and non-experts who are interested, besides representatives of public and private organizations improving their own awareness.

To conclude, the main originality of the proposal is to show a method to integrate traditional expert and volunteered glaciological data with glaciological data automatically extracted from incidental sources, so as to cope with the spatio-temporal scarcity of information about Alpine glaciers.

This work also discusses and proposes solutions to critical issues such as quality/reliability of data, their spatial and language features, protection and copyright, and web users' involvement.

Acknowledgments

Conflict of Interest

The authors declare no conflict of interest.

References

1. Lemke, P.; Ren, J.; Alley, R.B.; Allison, I.; Carrasco, J.; Flato, G.; Fujii, Y.; Kaser, G.; Mote, P.; Thomas, R.H.; *et al.* Observations: Changes in Snow, Ice and Frozen Ground. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; Solomon, S., Qin D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L., Eds.; Cambridge University Press: New York, NY, USA, 2007; pp. 338–383.

2. United Nations Environment Programme. *Global Glacier Changes: Facts and Figures*. Available online: http://www.grid.unep.ch/glaciers/ (accessed on 30 April 2013).

3. Haeberli, W.; Hoelzle, M.; Paul, F.; Zemp, M. Integrated monitoring of mountain glaciers as key indicators of global climate change: The European Alps. *Ann. Glaciol.* **2007**, *46*, 150–160.

4. Braithwaite, R.J. Glacier mass balance: The first 50 years of international monitoring. *Progr. Phys. Geogr.* **2002**, *26*, 76–95.

5. Paul, F.; Kääb, A.; Maisch, M.; Kellenberger, T.; Haeberli, W. Rapid disintegration of Alpine glaciers observed with satellite data. *Geophys. Res. Lett.* **2004**, *31*, doi: 10.1029/2004GL020816.

6. Haeberli, W.; Beniston, M. Climate change and its impacts on glaciers and permafrost in the Alps. *Ambio* **1998**, *27*, 258–265.

7. Zemp, M.; Haeberli, W.; Hoelzle, M.; Paul, F. Alpine glaciers to disappear within decades? *Geophys. Res. Lett.* **2006**, *33*, doi: 10.1029/2006GL026319.

8. Paul, F.; Kääb, A.; Haeberli, W. Recent glacier changes in the Alps observed by satellite: Consequences for future monitoring strategies. *Global Planet. Change* **2007**, *56*, 111–122.

9. Parr, T.W.; Ferretti, M.; Simpson, I.C.; Forsius, M.; Kovács-Láng, E. Towards a long-term integrated monitoring programme in Europe: Network design in theory and practice. *Environ. Monit. Assess.* **2002**, *78*, 253–290.

10. Bruns, A. Toward Produsage: Futures for User-Led Content Production. In Proceedings of Cultural Attitudes towards Communication and Technology, Tartu, Estonia, 28 June–1 July 2006.

11. Turner, A. *Introduction to Neogeography*; O'Really Media: Sebastopol, CA, USA, 2006.

12. Heipke, C.; Crowdsourcing geospatial data. *ISPRS J. Photogramm.* **2010**, *65*, 550–557.

13. *Community Collaborative Rain, Hail and Snow Network*. Available online: http://www.cocorahs.org/ (accessed on 30 April 2013).

14. *Alpine of the Americas Project*. Available online: http://alpineamericas.com/ (accessed on 30 April 2013).

15. *Applicazione Sperimentale per la Raccolta di Informazioni Legate Alla Condizioni di Innevamento in Piemonte*. Available online: http://remotesensing.arpa.piemonte.it/neve/ (accessed on 30 April 2013).

16. *SnowALP*. Available online: http://www.arpa.vda.it/ (accessed on 30 April 2013).

17. *Ghiacciai di un Volta*. Available online: http://www.ghiacciaidiunavolta.it/ (accessed on 30 April 2013).

18. Haklay, M.; Singleton, A; Parker, C. Web mapping 2.0: The neo-geography of the GeoWeb. *Geogr. Compass* **2008**, *22*, 2011–2039

19. Cohn, J.P. Citizen science: Can volunteers do real research? *BioScience* **2008**, *58*, 192–197.

20. Silvertown, J. A new dawn for citizen science. *Trend Ecol. Evolut.* **2009**, *24*, 467–471.

21. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.

22. European Parliament and Council of European Union. Directive INSPIRE. 2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Offic. J. Eur. Union* **2007**, 108, 1–14.

23. Flanagin, A.J.; Metzger, M.J. The credibility of volunteered geographic information. *GeoJournal* **2008**, *72*, 137–148.

24. Specia, L.; Motta, E. Integrating Folksonomies with the Semantic Web. In *The Semantic Web: Research and Applications*; Franconi, E., Kifer, M., May, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 624–639.

25. Ostermann, F.; Spinsanti, L. A Conceptual Workflow for Automatically Assessing the Quality of Volunteered Geographic Information for Crisis Management. In Proceedings of the 14th AGILE International Conference on Geographic Information Science, Utrecht, The Netherlands, 18–21 April 2011.

26. Smeulders, A.W.M.; Worring, M.; Santini, S.; Gupta, A.; Jain, R. Content-based image retrieval at the end of the early years. *IEEE Trans. Patt. Anal. Mach. Int.* **2000**, *22*, 1349–1380.

27. Wood, J. *Describing Place Through User Generated Content*. Available online: http://firstmonday.org/ojs/index.php/fm/article/view/3710/3035 (accessed on 30 April 2013).

28. Keßler, C.; Mau é, P.; Heuer, J.; Bartoschek, T. Bottom-Up Gazetteers: Learning from the Implicit Semantics of Geotags. In *GeoSpatial Semantics SE-6*; Janowicz, K., Raubal, M., Levashkin, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 83–102.

29. MacEachren, A.M.; Robinson, A.C.; Jaiswal, A.; Pezanowski, S.; Savelyev, A.; Blanford, J.; Mitra, P. Geo-Twitter Analytics: Applications in Crisis Management. In Proceedings of the 25th International Cartographic Conference, Paris, France, 3–8 July 2011.

30. Bordogna, G.; Criscuolo, L.; Carrara, P.; Pepe, M. A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. *Inform. Sciences* **2013**, in press.

31. Batini, C.; Scannapieco, M. *Data Quality: Concepts, Methodologies and Techniques*; Springer: Berlin, Germany, 2006.

32. Moellering, H. *A Draft Proposed Standard for Digital Cartographic Data*; National Committee for Digital Cartographic Data Standards: Columbus, OH, USA, 1987.

33. Chrisman, N.R. Issues and problems relating to cartographic data use, exchange and transfer: The role of quality information in the long-term functioning of a geographic information system. *Cartographica* **1984**, *21*, 79–88.

34. Popescu, A.; Grefenstette, G.; Bouamor, H. Mining a Multilingual Geographical Gazetteer from the Web. In Proceedings of IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Milan, Italy, 15–18 September 2009.

35. Kuhn, W. Ontologies in support of activities in geographical space. *Int. J. Geogr. Inf. Sci.* **2001**, *15*, 613–631.

36. Galloway, A.W.E.; Tudor, M.T.; Haegen, W.; Vander, M. The reliability of citizen science: A case study of Oregon White Oak stand surveys. *Wildlife Soc. Bull.* **2006**, *34*, 1425–1429.

37. *eBird Project*. Available online: http://ebird.org/content/ebird/ (accessed on 30 April 2013).

38. *Project Feeder Watch*. Available online: http://www.birds.cornell.edu/pfw/ (accessed on 30 April 2013).

39. Dickinson, J.L.; Zuckerberg, B.; Bonter, D.N. Citizen Science as an ecological research tool: Challenges and benefits. *Annu. Rev. Ecol. Evol. Syst.* **2010**, *41*, 149–172.

40. Hovland, C.I.; Irving, L.J.; Harold, H.K.; *Communication and Persuasion: Psychological Studies of Opinion Change*; Yale University Press: New Haven, CT, USA, 1953.

41. Metzger, M.J. Making sense of credibility on the Web: Models for evaluating online information and recommendation for future research. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 2078–2091.

42. Eysenbach, G.; Köhler, C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Brit. Med. J.* **2002**, *324*, 573–577.

43. Coleman, D.J.; Georgiadou, Y.; Labonte, J. Volunteered geographic information: The nature and motivation of produsers. *Int. J. Spat Data Infrastr. Res.* **2009**, *4*, 332–358.

44. Van Dijck, J. Users like you? Theorizing agency in user-generated content. *Media Cult. Soc.* **2009**, *31*, 41–58.

45. Huang, K.L.; Kanhere, S.S.; Hu, W. Are You Contributing Trustworthy Data? The Case for a Reputation System in Participatory Sensing. In Proceedings of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems, Bodrum, Turkey, 17–21 October 2010.

46. Reddy, S.; Shilton, K.; Burke, J.; Estrin, D.; Hansen, M.; Srivastava, M. Evaluating Participation and Performance in Participatory Sensing. In Proceedings of International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems, Raleigh, NC, USA, 4 November 2008.

47. De Longueville, B.; Luraschi, G; Smits, P.; Peedell, S; de Groeve, T. Citizens as sensors for natural hazards: A VGI integration workflow. *Geomatica* **2010**, *64*, 355–363.

48. De Solla, S.R.; Shirose, L.J.; Fernie, K.J.; Barrett, G.C.; Brousseau, C.S.; Bishop, C.A. Effects of sampling efforts and species detectability on volunteer based anuran monitoring programs. *Biol. Conserv.* **2005**, *121*, 585–94.

49. Genet, K.S.; Sargent, L.G. Evaluation of methods and data quality from a volunteer-based amphibian call survey. *Wildlife Soc. Bull.* **2003**, *31*, 703–14.

50. Lotz, A.; Allen, C.R. Observer bias in anuran call surveys. *J. Wildlife Manage.* **2007**, *71*, 675–79.

51. Fitzpatrick, M.C.; Preisser, E.L.; Ellison, A.M.; Elkinton, J.S. Observer bias and the detection of low-density populations. *Ecol. Appl.* **2009**, *19*, 1673–1679.

52. Delaney, D.G.; Sperling, C.D.; Adams, C.S.; Leung, B. Marine invasive species: Validation of citizen science and implications for national monitoring networks. *Biol. Invasions* **2007**, *10*, 117–128.

53. Antoniou, B.; Morley, J.; Haklay, M. Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. *Geomatica* **2010**, *64*, 99–110.

54. Friedland, G.; Choi, J. Semantic computing and privacy: A case study using inferred geo-location. *Int. J. Semant. Comput.* **2011**, *5*, 79–93.

55. Fisher, F. *Citizen, Experts, and the Environment: The Politics of Local Knowledge*; Duke University Press: Durham, NC, USA, 2000.

56. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environ. Plan. B-Plan Design* **2010**, *37*, 682–703.

57. Girres, J.F.; Touya, G. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* **2010**, *14*, 435–459.

58. Cipeluch, B.; Jacob, R.; Winstanley, A.; Mooney, P. Comparison of the Accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Leicester, UK, 20–23 July 2010.

59. Goodchild, M.F. Citizens as voluntary sensors: Spatial data infrastructure in the World of Web 2.0. *Int. J. Spat. Data Infrastr. Res.* **2007**, *2*, 24–32.

60. Connors, J.P.; Lei, S.; Kelly, M. Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Ann. Assn. Amer. Geogr.* **2012**, *102*, 1267–1289.

61. Elwood, S.; Goodchild, M.F.; Sui, D.Z. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Ann. Assn. Amer. Geogr.* **2012**, *102*, 571–590.

62. Ferster, C.J.; Coops, N.C. A review of earth observation using mobile personal communication devices. *Comput. Geosci.* **2013**, *51*, 339–349.

63. Lane, N.; Miluzzo, E.; Lu, H.; Peebles, D.; Choudhury, T.; Campbell, A. A survey of mobile phone sensing. *IEEE Commun. Mag.* **2010**, *48*, 140–150.

64. De Cristofaro, E.; Soriente, C. PEPSI : Privacy-Enhanced Participatory Sensing Infrastructure. In Proceedings of the Fourth ACM Conference on Wireless Network Security ACM, Hamburg, Germany, 14–17 June 2011.

65. Ganti, R.K.; Pham, N.; Tsai, Y.E.; Abdelzaher, T.F. Pool View: Stream Privacy for Grassroots Participatory Sensing. In Proceedings of the 6th ACM Conference on Embedded Network Sensor System, Raleigh, NC, USA, 4–7 November 2008.

66. *EnvEurope Project*. Available online: http://www.enveurope.eu/ (accessed on 30 April 2013).

67. *NERC Data Policy—Guidance Notes*. Available online: http://www.nerc.ac.uk/research/ sites/data/documents/datapolicy-guidance.pdf (accessed on 30 April 2013).

68. *LATHUILE FLUXNET DATA—Data policy and Terms of Reference*. Available online: http://fluxnet.ornl.gov/sites/default/files/Policy_LaThuile_Final.pdf (accessed on 30 April 2013).

69. European Parliament and Council of the European Union. European Parliament and Council Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Offic. J. Eur. Union L* **1995**, *281*, 31–50.

70. European Parliament and Council of the European Union. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector. *Offic. J. Eur. Union L* **2002**, *201*, 37–47.

71. European Parliament and Council of the European Union. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. *Offic. J. Eur. Union L* **2001**, *167*, 10–19.

72. Rashid, A.M.; Ling, K.; Tassone, R.D.; Resnick, P.; Kraut, R.; Riedl, J. Motivating Participation by Displaying the Value of Contribution. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, QC, Canada, 22–27 April 2006.

73. Baldocchi, D.; Falge, E.; Gu, L.; Olson, R.; Hollinger, D.; Running, S.; Anthoni, P.; Bernhofer, C.; Davis, K.; Evans, R.; *et al.* FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Amer. Meteorol. Soc.* **2001**, *82*, 2415–2434.