Czech University of Life Sciences Prague

Bachelor’s Thesis

2022 Ivana Kazazovic

Czech University of Life Sciences Prague

Faculty of Environmental Science

Open data sources on natural disasters and climatological extremes

Bachelor’s Thesis

Author: Ivana Kazazovic
Supervisor: prof. Ing. Martin Hanel, PhD
Year of publication: 2022

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

BACHELOR THESIS ASSIGNMENT

MA. Ivana Kazazovic, MA

Environmental Data Science
Ecology

Thesis title

Open data sources on natural disasters and climatological extremes

---

Objectives of thesis

1. Identify the most valuable online sources of open data in the domains of ecology, natural disasters and climatological extremes

2. Evaluate differences in how data is collected, organized and presented

3. Compare the availability of data in terms of type, geographical and temporal ranges

4. Examine potentials and limitations in the available data products

Methodology

Part 1: Literature research

1. Identify the most recent and high-quality available research on data collection in the ecological domain to get an understanding of available methods and technological limitations

2. Establish the actual as well as future potential of open data's role in mitigating natural disasters and increasingly frequent weather extremes

Part 2: Practical evaluation

1. Compare the characteristics of available data between sources

3. Document use cases utilizing open data on natural disasters and climatological extremes

3. Assess opportunities and limitations related to data sources

The proposed extent of the thesis

40 pages

Keywords

risk assessment; disaster monitoring; open data

Recommended information sources

Hao, Z., Yuan, X., Xia, Y., Hao, F., & Singh, V. P. (2017). An Overview of Drought Monitoring and Prediction Systems at Regional and Global Scales. Bulletin of the American Meteorological Society, 98(9), 1879–1896. doi:10.1175/bams-d-15-00149.1

Hawken, S., Han, H., & Pettit, C. (Eds.). (2020). Open Cities | Open Data. doi:10.1007/978-981-13-6605-5

Michael C. Dietze, Ecological forecasting, Princeton University Press, 2017.

Mohamed Gad-el-Hak, Large-Scale Disasters: Prediction, Control, Mitigation, Cambridge University Press, 2008.

Radocaj, D., Obhodas, J., Jurisic M., Gasparovic M. (2020). Global Open Data Remote Sensing Missions for Land Monitoring and Conservation: An Overview. Land. doi:10.3390/land9110402.

Rocchini, D., Petras, V., Petrasova, A., Horning, N., Furtkevicova, L., Neteler, M., ... Wegmann, M. (2017). Open data and open source for remote sensing training in ecology. Ecological Informatics, 40, 57–61. doi:10.1016/j.ecoinf.2017.05.004

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., ... Zwiers, F. W. (2017). Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. Weather and Climate Extremes, 18, 65–74. doi:10.1016/j.wace.2017.10.003

Expected date of thesis defence

2021/22 SS – FES

The Bachelor Thesis Supervisor

prof. Ing. Martin Hanel, Ph.D.

Supervising department

Department of Water Resources and Environmental Modeling

Advisor of thesis

Ing. Lukáš Tůma

Electronic approval: 24. 3. 2022

prof. Ing. Martin Hanel, Ph.D.

Head of department

Electronic approval: 25. 3. 2022

prof. RNDr. Vladimír Bejček, CSc.

Dean

Prague on 29. 03. 2022

Author’s Statement

I hereby declare that I have independently elaborated the bachelor/final thesis with the topic of: Open data sources on natural disasters and climatological extremes, and that I have cited all of the information sources that I used in the thesis as listed at the end of the thesis in the list of used information sources.

I am aware that my bachelor/final thesis is subject to Act No. 121/2000 Coll., on copyright, on rights related to copyright and on amendments of certain acts, as amended by later regulations, particularly the provisions of Section 35(3) of the act on the use of the thesis. I am aware that by submitting the bachelor/final thesis I agree with its publication under Act No. 111/1998 Coll., on universities and on the change and amendments of certain acts, as amended, regardless of the result of its defence.

With my own signature, I also declare that the electronic version is identical to the printed version and the data stated in the thesis has been processed in relation to the GDPR.

Date: 23.03.2022
Place: Prague, Czech Republic

Signature:

Abstract (English)

The field of data science, wherein one aspect includes large amounts of data being used for creating simulations, has been growing in the field of environmental science in recent years. Modelling of the environment has been used to tackle issues related to climate change, pollution, deforestation, and many other pressing ecological issues. Access to open data has also been growing, with an increasing number of online platforms providing large amounts of information for free on the web. Considering that natural disasters, in particular those related to climatological extremes in precipitation and temperature such as floods and droughts, have been becoming more frequent in recent years, a lot of the data collected and published is related to these risks. While it would be beyond the scope of this thesis to create a detailed catalogue of all providers of environmental data online, ten platforms were chosen for evaluation in relation to their potential to provide valuable insights in light of these threats. The data providers were chosen based on providing a unique perspective or service which makes use of environmental data, in an attempt to showcase different possibilities, as well as limitations of models of data providers.

Abstract (Czech)

Význam vědy o datech (data science) v oblasti věd o životním prostředí v posledních letech rychle roste zejména v souvislosti s modelováním a simulací environmentálních problémů. Modelování životního prostředí se používá k řešení problémů souvisejících se změnou klimatu, znečištěním, odlesňováním a mnoha dalšími naléhavými ekologickými problémy. Zvyšuje se take dostupnost otevřených datům, přičemž stále více online platforem poskytuje velké množství informací zdarma na webu. Vzhledem k tomu, že přírodní katastrofy, zejména ty, které souvisejí s klimatologickými extrémy ve srážkách a teplotách, jako jsou povodně a sucha, jsou v posledních letech stále častější, řada shromážděných a publikovaných dat souvisí s těmito riziky. I když by bylo nad rámec této práce vytvořit podrobný katalog všech poskytovatelů environmentálních dat online, bylo pro hodnocení vybráno deset platforem ve vztahu k jejich potenciálu poskytovat cenné poznatky ve světle těchto hrozeb. Poskytovatelé dat byli vybráni s cílem demonstrovat jedinečné perspektivy nebo služby využívající environmentální data, ve snaze ukázat různé možnosti a také související omezení.

Keywords (English): Spatial data, time series data, climate modelling, natural hazards, flood

Keywords (Czech): Prostorová data, časov řady, klimatické modelování, přírodní rizika, povodně

Table of Contents

1. Introduction …………………………………………………………………1
2. Objectives ………………………………………………………………… 3
3. Literature review ………………………………………………………… 4
3.1 Defining terms ………………………………………………………… 5
3.1.1 Natural disasters and climatological extremes ……………………… 5
3.1.2 Big data and open access ……………………………………………8
3.2 Environmental data ……………………………………………………… 9
3.2.1 Types of environmental data ………………………………………… 9
3.2.2 Data in the context of natural disasters …………………………… 11
3.2.3 Open data in environmental research ……………………………… 10
3.3 Modelling ……………………………………………………………… 14
3.3.1 Research into modelling within the environmental science sector ……………………………………………………………………14
3.3.2 Types of models ……………………………………………………… 15
3.3.3 Limitations …………………………………………………………… 16
3.4 Room for future research ……………………………………………… 17
4. Evaluation ………………………………………………………………… 17
4.1 Methodology …………………………………………………………… 17
4.2 Data projects with organisations or individuals ……………………… 18
4.2.1 Open Data For Resilience Initiative (OpenDRI)............................... 19
4.2.2 Digital Green ………………………………………………………… 21
4.2.3 EO4SD ……………………………………………………………… 21
4.3 Data made available directly ………………………………………… 23
4.3.1 Water Data Portal (WDP) ……………………………………………23
4.3.2 Global Agricultural Research Data Innovation & Acceleration Network (GARDIAN) ………………………………………………… 26
4.3.3 Digital Earth Africa ………………………………………………… 28
4.3.4 Data Bank ……………………………………………………………31
4.3.5 World Environment Situation Room ……………………………… 33
4.3.6 European Environment Agency …………………………………… 36
4.3.7 The Copernicus Programme …………………………………………38
4.3.8 Data.gov ……………………………………………………………… 41
4.3.9 AI for Earth ………………………………………………………… 43
5. Discussion ………………………………………………………………… 44
5.1 Providing data through agreements and providing data openly on the web ……………………………………………………………………………44
5.1.1 NGO sector ……………………………………………………………45
5.1.2 Governmental sector …………………………………………………45
5.1.3 Industry ………………………………………………………………46
5.2 Needs of researchers …………………………………………………46
5.3 Natural hazards, climatological extremes …………………………46
5.4 Modelling ……………………………………………………………46
6. Conclusion …………………………………………………………………47
7. Bibliography ………………………………………………………………49
8. List of figures ……………………………………………………………… 54

Appendix A ……………………………………………………………………………………….. 54
Appendix B …………………………………………………………………………………………54
Appendix C …………………………………………………………………………………………54

1. Introduction

The ongoing boom in large scale data collection, driven by rapid improvements in technology, has revolutionised decision making across industries, including environmental science. Access to information has the ability to influence and optimise our understanding of a problem as well as its possible consequences and solutions with a higher degree of precision and in less time than was previously possible (Brynjolfsson, 2016).

When it comes to documenting the physical environment, there are many different types of data that can be collected. One major type is discrete time series data where environmental variables are measured through on-location stations or surveys, an important component in understanding environmental trends through time (Brbulescu, 2018). Having high quality data of this type helps us make predictions of where things are headed, by analysing and understanding the previous state of the variables being measured. Most commonly these include precipitation and greenhouse gases, but they can also measure many other things such as pollutants and other chemicals (Gouveia, 2000), species counts for biodiversity tracking (Staudhammer, Escobedo and Blood, 2018), how much electricity is being used in an area (Jain, Quamer and Pamula, 2018), etc.

The other common type of data used in environmental science is spatial data, which is collected through airborne technology such as satellites, drones and planes. This type of data is useful in observing the physical characteristics of the Earth, which combined with various equipment such as sensors and measurements provides valuable information and helps us solve different kinds of problems like monitoring deforestation (Hadi et al., 2018), observing inaccessible places (Avisse et al., 2017), tracking the state of vegetation (Wozniak et al., 2020), and much more (Bernstein, 2020).

The information that these types of data provide, therefore, have the potential to aid in everything from making existing industries like agriculture operate more efficiently, to more precise predictions of natural disasters which could save lives by improving our immediate response, as well as deepening our understanding of our environment and the impacts we have on it so that we can mitigate issues like climate change and pollution (Sebestyén, 2021).

Open data, in particular, is a powerful potential source of change as it makes information freely available (Roche, 2020). This accessibility makes it easier to achieve worldwide cooperation across different industries, including governments, the NGO sector, academic and research faculties as well as invested communities and individuals. Hence it is made possible for insights and ideas to be shared between groups which might have had a harder time collaborating in the past. Furthermore, regions of the world with fewer resources to create their own monitoring systems benefit due to increased access to valuable information about their ecology, improving their abilities to manage their resources and prepare for disasters, to which they are often the most vulnerable.

However, ecological systems are notoriously complex and difficult to predict. As Michael Dieze put it in his book *Ecological forecasting*: “data are noisy, and dynamics can be idiosyncratic, varying from system to system or site to site in ways that defy our current understanding” (2017). He continues to argue that if we are to make climate modelling a more accurately predictive science, we have to both enhance our understanding of the mathematical component of prediction, namely probability and uncertainty, as well as have a deep and clear understanding of the technology that collected the data in the first place, so we can account for any idiosyncrasies that may be present. Therefore, despite advancements in technology there is still work that needs to be done to make it possible to use these sources of information effectively and accurately.

For this thesis, I will take a closer look at the available literature and research on open access data as well as ecological modelling. Alongside this, I will provide an in-depth analysis of a selection of online sources of environmental data to gain a better understanding of what role they play and what their potential, goals and limitations are in predicting, understanding and mitigating environmental disasters and ecological extremes. This thesis is inspired by a collaboration through an internship with the Prague based company Big Terra, which specialises in using data for optimising farming practices across the world. They have provided five sources of open data that they use in their work which I will be including in this analysis, alongside seven more which through individual research were deemed appropriate for reasons which will be detailed in the methodology section.

2. Objectives

The objective for this thesis is a systematic review of open data sources and their potential for the purposes of predicting and mitigating natural disasters and climatological extremes. Both actualised as well as future potential will be evaluated, and the sources individually presented and compared to showcase different methodologies in attaining and presenting data products. This can be further split into four categories:

2.1. Identify the most valuable online sources of open data in the domains of ecology, natural disasters and climatological extremes

Due to a large number of available sources of ecological data on the internet, criteria for selecting sources to include in this research need to be defined. Each source should provide a unique contribution and a deeper understanding of the data landscape and which problems they aim to solve.

2.2. Evaluate differences in how data is collected, organised and presented

Each selected source will be assessed to identify the technologies used to collect the presented data, as well as the user interface (UI) of their websites for clarity and ease of navigation and data download. The main features of the data such as temporal and geographical range, availability of metadata and diversity of data types will be presented, analysed and compared.

2.3. Review of existing research

A deep dive into the existing research in the domains of ecological forecasting and open data will be presented. This will showcase the current understanding of the potentials and limitations of this field and will provide the basis off of which the evaluation of the sources will be conducted.

2.4. Examine potentials and limitations in the available data products

I will critically assess each data source's actualised and future potential for impact through environmental data. Their individual strengths and weaknesses will be assessed and compared to provide a comprehensive overview of the environmental open access data landscape.

3. Literature Review

With the effects of climate change becoming increasingly apparent over time, environmental modelling has the potential to increase our understanding of physical phenomena and the state of the environment, by extension creating the potential to predict and mitigate future disasters. This ability to create predictions of climate extremes has therefore been identified as a major area which needs further progress in the field of climate research (Sillmann et al., 2017).

However, the environmental systems which are at play to determine the climate are notoriously hard to predict. Weather systems are the basis of which chaos theory, the study of unpredictable, seemingly random behaviour, is founded on (Gad-ElHak, 2009). By their very nature, these dynamic processes are a challenge to model due to being influenced by many different elements, the relationships between which are not yet fully understood.

This is why researchers are actively working on perfecting our ability to model these processes. Despite the complexity involved and a degree of uncertainty where extremes are concerned, most environmental processes are driven by a small subset of interactions and processes (Dietze, 2017). In part, the ability of researchers to advance our knowledge in this field depends on the quality and quantity of the data that is available for them to study. Another aspect is the advancement of technological potential and mathematical understanding needed to increase the degree of accuracy in the models.

In this review I will highlight the most relevant research that has been done in recent years in the field of environmental data collection and forecasting, and showcase the potential as well as limitations which exist today.

3.1 Defining terms

3.1.1 Natural disasters and climatological extremes

Natural disasters are defined as large-scale events which cause devastation and are triggered by natural processes. They are not defined strictly by their cause but by their impact on loss of life and property to the community that exists within the area they hit. "It is generally accepted among environmental geographers that there is no such thing as a natural disaster. In every phase and aspect of a disaster— causes, vulnerability, preparedness, results and response, and reconstruction—the contours of disaster and the difference between who lives and who dies is to a greater or lesser extent a social calculus" (Gad-ElHak, 2009). It is therefore the impact on human communities that defines these events as disasters.

Some disasters are independent of the climate and are related to other natural processes, such as earthquakes and volcanic activity. Others however are connected to the weather, such as droughts and floods which are directly related to climatological extremes in precipitation. A climatological extreme is defined as a weather event which is rare within the context of its statistical reference, meaning that it is within the tenth or ninetieth percentile for the region (Gad-ElHak, 2009). In this sense, the definition of a climatological extreme will vary from place to place as different parts of the planet have varying baselines for weather.

They can be categorised according to severity, which can be defined for instance by human beings injured or total area affected (Gad-ElHak, 2009). This division is shown in Figure 1 below:

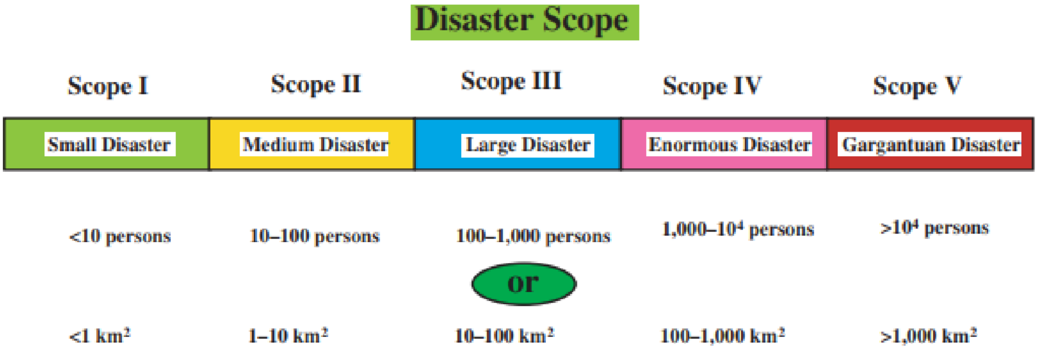


Fig 1. Classification of disaster severity (Gad-ElHak, 2009)

Within this system, if a single event causes a mismatch in the scope of impact on area and human beings injured, the larger of the two is taken into account for categorisation purposes. For example, an event which injures fewer than ten people, but damages an area that is greater than one km squared but smaller than ten, would be classified as a Medium disaster.

The choices human beings made in the initiation of these events can have a profound impact on the severity of the event. Access to accurate, timely information has increased the ability of governments and emergency response teams to save lives and make a difference in the wake of disasters. Early warning systems have played a crucial role in reducing the deaths from weather, climate and water hazards (IPCC, 2012). This trend can be observed in Figure 2 below, showcasing the loss of life from natural disasters over time:

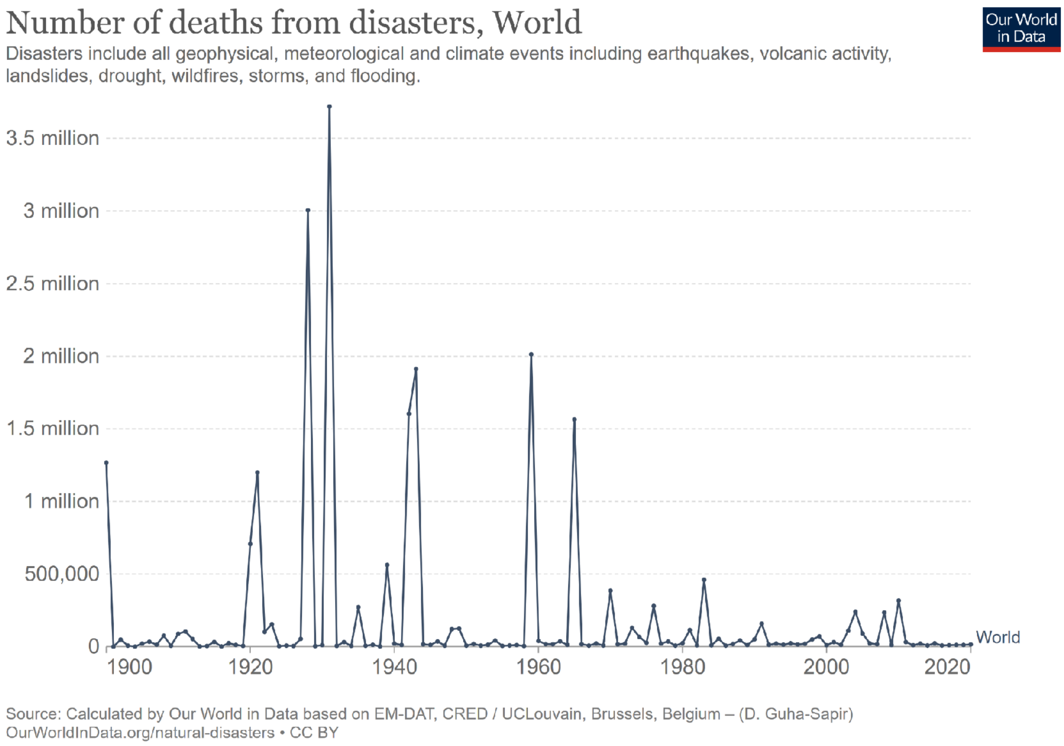


Fig 2. Note. From *Deaths by natural disasters* by Ritchie H. and Roser M., 2014 (https://ourworldindata.org/natural-disasters#citation)

Furthermore, if we take a closer look at the numbers for each type of disaster in Figure 3, we can notice that in terms of economic damages, there has been an increase in the past decade compared to the previous with regards to damages incurred from flooding, drought and wildfires.

On average in the 21st century, the rate of both droughts and floods have increased as a higher level of precipitation falls seasonally, and summers become drier (Christidis and Stott, 2021). Storms, earthquakes and floods have been the most costly natural hazards in terms of GDP in recent decades.

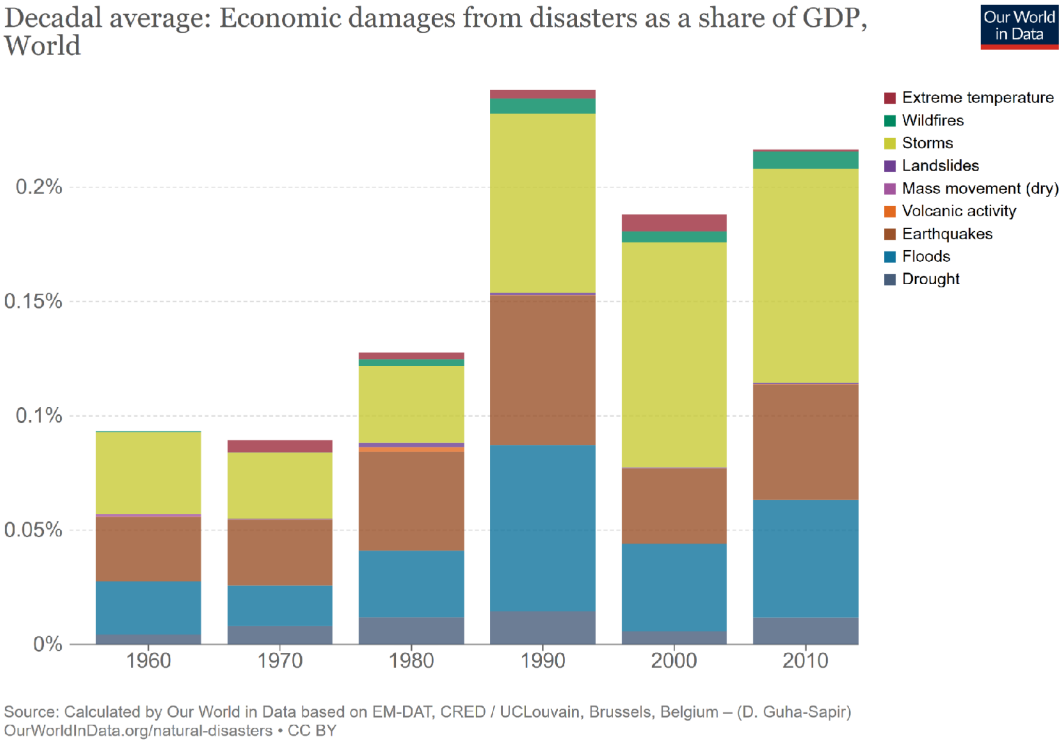


Fig 3. Note. From *Economic damages from disasters* by Ritchie H. and Roser M., 2014 (https://ourworldindata.org/natural-disasters#citation)

However modelling climate extremes becomes more difficult as we try to increase the window of time before the event. Generally, short term forecasts have a higher degree of accuracy, as the accuracy of the prediction decreases as the distance in between the time of the prediction and the event increases (Ren et al., 2021).

In this way, the severity of the impact is closely reliant on the preparedness and response time of authorities and emergency services, as well as the affected communities and individuals within them. Access to high quality data regarding these events is therefore central in our ability to mitigate the vulnerability of communities to natural disasters, and increasing the accuracy of the data and models is crucial in predicting and mitigating the impacts of extreme weather and other natural hazards by giving communities extra time to prepare for the effects of drought or extreme precipitation.

3.1.2 Big data and open access

There is currently no single rigorous definition of big data (Mayer-Schönberger and Cukier, 2013). While on the one hand the large volume of raw information has opened up doors for finding insights that were before not accessible, the ability to manage it meaningfully and efficiently has posed its own challenges as the volume of information collected has grown through improvements in technology. Tools have had to be developed which would allow the sorting and analysing of the newly available material.

Data sharing policies, on the other hand, refer to the negotiated agreements within and in between organisations for the exchange of information. While some agencies may restrict access to data due to the nature of the information being sensitive or for competitive reasons, in other cases the raw information may be provided freely to all users in an unrestricted manner in what is known as open access data (Borowitz, 2017). Given the wide range of access policies which are available, this thesis will mainly focus on open access data as it is made available to the general public freely, with only three exceptions to showcase particular organisations which use data in a more specialised manner.

3.2 Environmental data

3.2.1 Types of environmental data

One of the key features in the field of data science specialising in the environment is that there are many different types of data which are collected in different ways, lending to a high degree of heterogeneity. A key challenge is combining the different data types to gain deeper and more accurate insights about the environment, as well as the diverse methodologies that researchers might use across the world in different institutions which might not be easily compatible with each other for collaboration. Hence a key challenge in the field is developing and integrating standardised mechanisms and techniques for working with big data.

Data types in this field can be divided into the following main categories (Blair et al., 2019):

1. Remote sensing data which documents the planet from a distance without direct contact, usually through satellites and other aircraft. These can be further split into passive sensing such as photography (Morgan et al, 2010), and active sensing such as RADAR where radio waves are employed (Zhang

et al, 2021), or LIDAR where lasers are used to measure time travelled from the Earth's surface (Harpold, 2015). This type of data has been becoming increasingly available openly through government agencies such as the EU Copernicus programme or NASA's LandSat archives making it possible for individuals and organisations to make use of this type of information.

2. Sensors which monitor localised phenomena in closer proximity. Common applications for this type of data include weather monitoring stations, which record rates of and changes in precipitation and temperature, as well as pollution monitoring for keeping track of chemicals in the water or air. This type of data collection has the potential to grow even further in the near future as more devices, through the introduction of The Internet of Things technology which include various sensors, are deployed more widely allowing for wide-ranging large volumes to be collected (Hart & Martinez, 2015).
3. Surveys or field campaigns of individuals or groups which collect data directly, for example in the fields of biodiversity and soil quality. These can include both specialists working in a particular field, as well as citizen science meaning crowdsourcing, those which are personally interested in the field and make use of personal resources for the collection of this data, allowing large volumes of data to be collected and providing a significant contribution to research (de Sherbinin et al., 2021).
4. Historical records which, when available, provide a valuable baseline of comparison to previous norms and trends (Maliva, 2021). This type of data is extremely valuable in the field of modelling and the digitalisation and care of historical records, particularly on a local level, is of high priority as the lack of access to this type of data due to slow progress in the area of digitalisation has made much of this data inaccessible to researchers (Easterday et al., 2018).
5. Outputs from previous models which were generated and then stored also provide valuable information and can help create a more complete picture with further modelling work. The standardisation of methodologies and technologies used is valuable in this area particularly as it would allow the reuse of models and therefore a higher degree of efficiency, rather than

researchers having to remake them from scratch each time they are needed (Holzworth, Huth and de Voil, 2010).

6. Data mining of internet platforms, to make use of the vast amounts of information that is uploaded on the web regularly. Examples of these uses include studying particular phenomena such as water levels through images uploaded on the web (Cervone et al., 2015), as well as using social media to create early warning systems for flooding (Restrepo-Estrada et al., 2018).

3.2.2 Data in the context of natural disasters

The use case for each data type varies as they make it possible to address different types of problems. In the field of using data in particular as it relates to natural disasters, the objectives can be split into three broad categories (Goswami et al., 2018):

1. The prediction of natural disasters in terms of time, place and magnitude. For example, through the use of machine learning models to identify dangers of floods by studying the historical records of the physical environment and optimising early detection in any changes which might signal an incoming natural hazard (Maspo et al., 2020).
2. Improving detection of events. Detection of natural disasters can also be done through the monitoring of social media websites, wherein data science techniques are needed to sort through massive amounts of noisy data as users are actively responding to an event in their environment (Said et al., 2019)
3. The management of the disaster by authorities can be improved by having access to clear and accurate data about the impacts incurred. Social media data has been shown to be valuable in the field of disaster response, for example through more efficient detection of sub-events such as collapsing of buildings, broken pipes, and more (Belcastro et al., 2021).

3.2.3 Open data in environmental research

Beyond government agencies and individual interests, access to environmental data plays an active role in the ability of researchers to conduct their work and advance their fields. A survey conducted in 2016 assessed various facets of data accessibility and how significant each component of its availability is to members of the scientific community across 80 countries. Their findings showed a high degree of emphasis on the importance of open availability of data, as researchers, data scientists, data managers and technologists from the environmental, earth, marine and polar science fields that participated in the survey highlighted through their responses (Schmidt et al., 2016).

While the ability to find well defined quality information and metadata, or information about the data, were rated as the highest and second highest priority respectively, ease of access is considered to be most important by 76% of responders showcasing the priority that freely accessible information has in the field of research. This is shown in Figure 3 below:

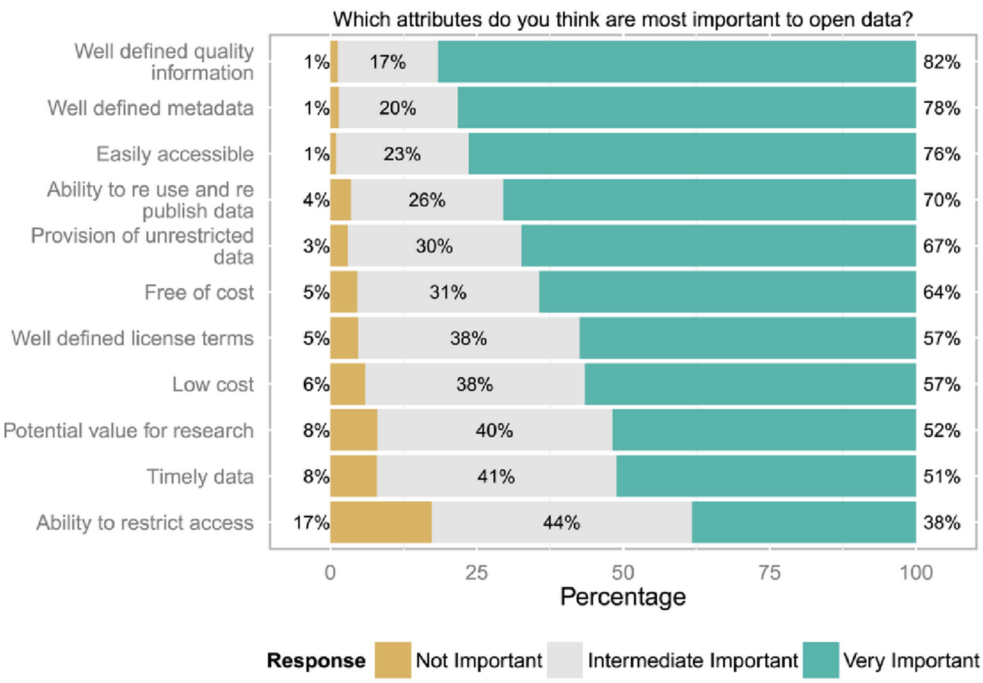


Fig 4. Which attributes are most important to open data (Schmidt et al., 2016). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695 (14.03.2022)

Furthermore, within the same paper, four out of five people surveyed claimed that open data is necessary for advancing research within their field or community, highlighting the role that ease of access to raw information plays in the advancement of knowledge in these fields (Figure 5).

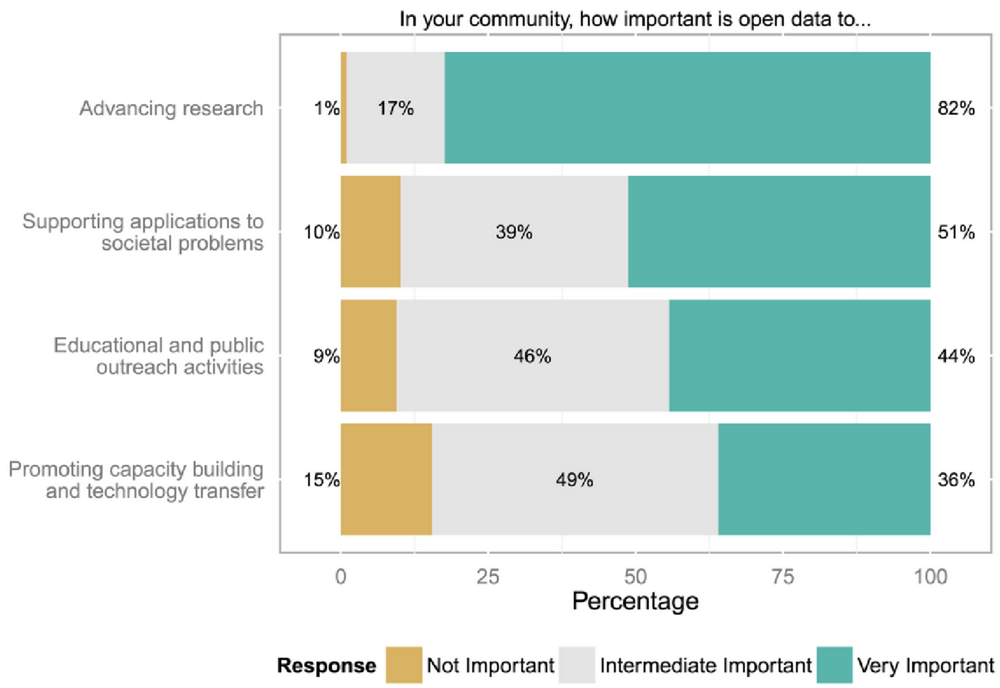


Fig 5. How important is open data in your community (Schmidt et al., 2016) https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695 (14.03.2022)

Finally, Figure 6 shows that regarding the resources that researchers use to find relevant data for their research, the main ones were journals and the web.

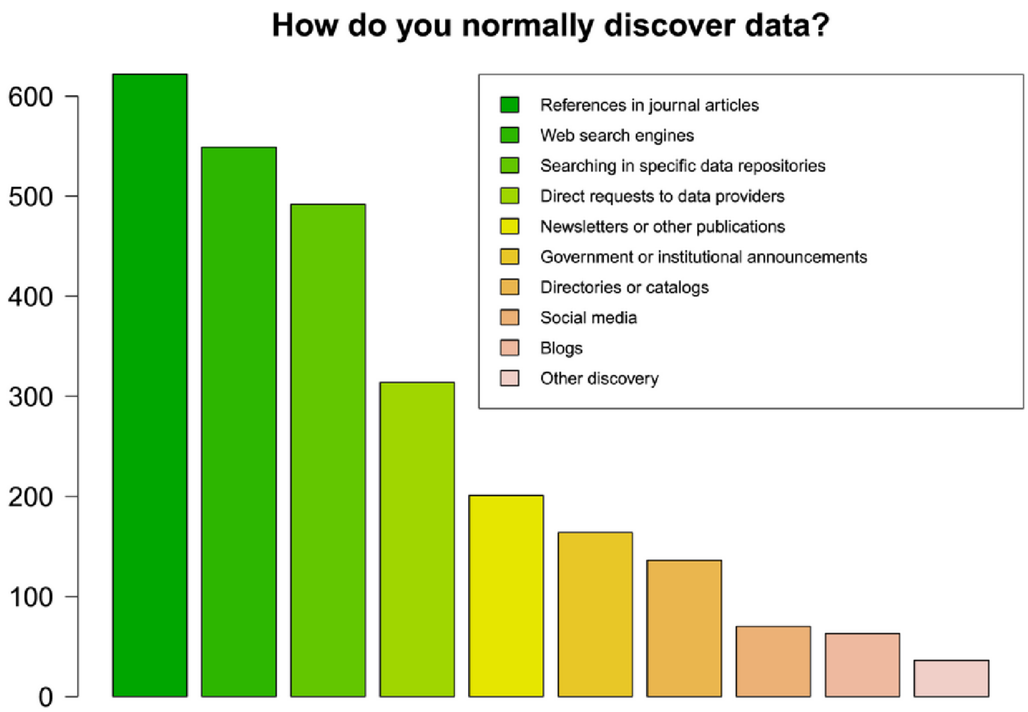


Fig 6. How do you discover data (Schmidt et al., 2016) https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695 (14.03.2022)

This research showcases how important having free and open access to data is to the advancement of research in environmental science, and hence our understanding of the physical environment and improving our ability to create accurate understandings and models of phenomena. Furthermore, given that the web plays a crucial role in the aggregation and discovery of data, the need for a methodological review of available sources is highlighted.

3.3 Modelling

The recreation of environmental processes through the use of models has been a key component in the research field as it has advanced our ability to understand and predict natural disasters and climatological extremes, as well as many other components of our environment (Duan, 2020).

3.3.1 Research into modelling within the environmental science sector

The graph below, Figure 7, showcases the numbers of research papers published, as found through Google Scholar, containing the keywords ecology, natural disaster or climate change combined with modelling (adjusted to account for both the American English and British English spellings of the word. Raw data is shown in Appendix A).

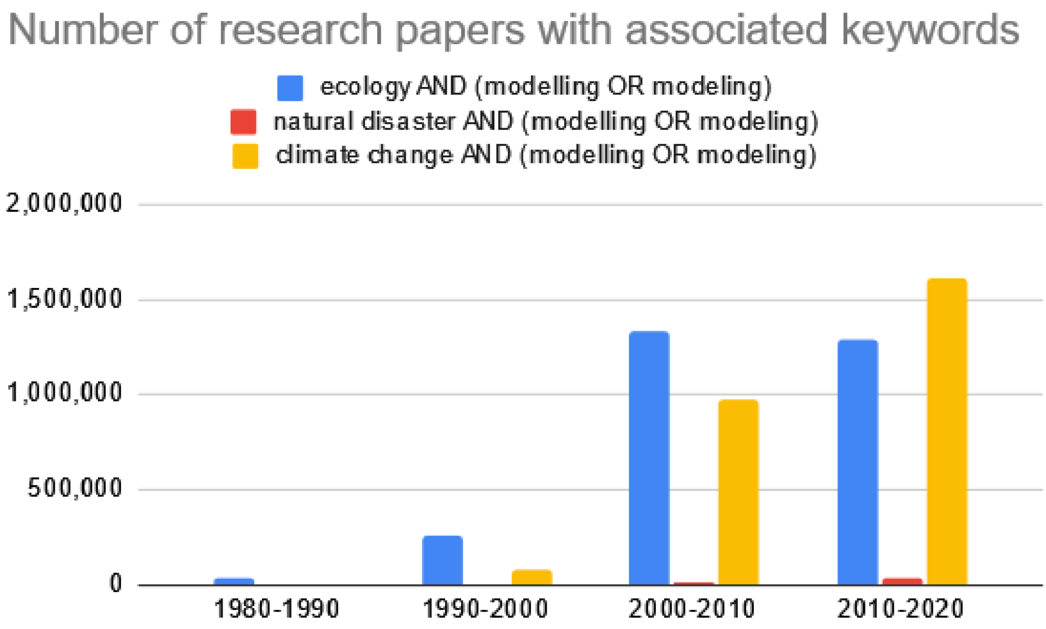


Fig 7. Number of research papers with the relevant keywords

Taking into consideration certain limitations and uncertainties (Refsgaard, 2007), such as a lack of full understanding of certain processes, modelling has been getting more accurate and widespread as its applications are utilised in an increasing number of fields. However, climate models may not be equally representative of reality depending on the region they are simulating due to differences in availability of the data (Sillmann et al, 2017). Improving the theories and hierarchies of models is here pointed out as the key challenge which would lead to an increase in accuracy in modelling.

3.3.2 Types of models

Models can be split into two categories (Blair et al, 2019):

1. Process models which are meant to capture the underlying physical processes being considered (van der Meer, 2013)
2. Data-driven models which are derived from past, possibly more complex models or are based on statistical fits to observed data (Knüsel, 2020)

Model simulations are then used to try and understand uncertainties and sensitivities within the system. Furthermore, global climate models fall within the category of integrated modelling systems (Weyant, 2017), due to many different aspects of the environment being modelled simultaneously where each process is individually simulated. The accurate reproduction of these systems therefore involve some challenges related to data science, including the accurate interpretation and propagation of uncertainties, as well as accounting for the complexity of the relationships between the phenomena and feedback loops.

Furthermore, while historical records can be used to create simulations of likely future scenarios, the past is not always indicative of the future. This is only becoming increasingly relevant as the impacts of climate change bring about more frequent extremes in the weather. As such, a solution in the form of data assimilation, or combining models with observations, has been brought forward as a solution (Lahoz et al, 2010). This method is meant to bring together the theory of how systems work through the use of established numerical models with the observed phenomena, providing the highest degree of accuracy in forecasting.

Finally, researchers across the globe do not currently have a unified methodology or technology which is used for modelling, meaning that much of the work being done on the field is difficult to assimilate and therefore build off of due to mismatches in computational software or programming languages being used. The authors point to the potential of Cloud Computing to address this issue, as making simulations accessible in this way can improve the workflow by making it more efficient and facilitate further research. Additionally, these services allow for the storage and management of large amounts of data which would allow researchers to make better use of these computational resources.

3.3.3 Limitations

As discussed in previous chapters, the main limitations within this field are in addressing the complexity of the systems that are being modelled, and assimilating knowledge and previous research into a cohesive whole so as to advance the field more efficiently.

Making data openly available provides access to invaluable information but it also creates an additional problem of sorting through the material and finding what is needed, while also making sure that the data collected was done so in a reliable and repeatable way. A need for an easy to use system of sorting through data has presented itself, as it can be scattered and hard to find through individual research.

3.4 Room for future research

Further research is necessary to reach a higher degree of accuracy and understanding using modelling tools in this sector. More localised projects which specialise on particular technologies and geographical areas would be valuable for extracting insights about which methodologies work best in the long run and for different problems, and how different components of the data collection process play out in practice. The potential for individuals and smaller organisations to make use of open access data, and which elements are most valuable for these users should also be assessed, to see if there is potential to provide greater support leading up to and during times of crisis.

4. Evaluation

In this chapter a detailed overview of each chosen source of online data will be provided.

4.1 Methodology

The first step for this part of the thesis was to choose the data sources which will be analysed. A total of twelve were chosen. OpenDRI, Water Data Portal, GARDIAN, Digital Green and Digital Earth Africa were provided by Big Terra through the collaborative internship. These are sources of data which they either use in their work with clients to analyse agricultural data directly, or they highlighted as interesting examples of data providers within the industry. Data.gov, Data Bank, World Environment Situation Room, European Environmental Agency, the Copernicus Program, EO4SD and AI for Earth were found through independent research to be large and diverse data providers.

Each of these providers has a different focus with the data they provide, the approach they take to presenting and using the data and overall goals and problems they choose to tackle. The differences in these focuses and presentation will be further discussed in greater detail.

To showcase the differences in a meaningful way, examples of representative use cases from each data source will be presented. The nature of this example will vary depending on the source in question, as they have significant differences and therefore are difficult to compare in a direct way.

These sources could be organised in several different ways. They could be split into type of data, geographic region which is covered or the sources and technologies they use. Alternatively, they could be split by organisational type, as some of these sources are private companies, while others are part of the NGO network and some are government agencies. For this thesis I will be using a two-way split. Firstly I will separate OpenDRI, Digital Green and EO4SD as they are not raw data providers but work closely with individuals and organisations using data to solve environmental problems, and therefore showcase one type of model for working with environmental data. The others are all direct providers of open access data and will be divided into the private, NGO and government sectors.

Background information regarding the platforms will also be presented. Information about their mission statements, funding and affiliations will provide a clearer picture about the data providers and allow for a better understanding of their potential strengths and weaknesses.

4.2 Data projects with organisations or individuals

In this chapter OpenDRI, DigitalGreen and EO4SD’s data services will be presented, which differ from the others in that rather than providing open access data directly they use data on specific projects to address highly specialised problems regionally, through collaboration with the local community.

The datasets will be presented in three parts: the introduction will provide background information about the platform, the features of the data will be briefly explained, and lastly an example of a use case will be showcased. Although the nature of the data and projects vary greatly, this will provide a standardised methodology by which to analyse and compare them.

4.2.1 Open Data For Resilience Initiative (OpenDRI)

Available at: https://www.opendri.org/

I. Introduction to OpenDRI

Created under the Global Facility for Disaster Reduction and Recovery (GFDRR) of the World Bank, the initiative's mission statement is to combat the impacts of natural hazards on vulnerable communities through the use of open data. Founded in 2011, they work directly with governments to share, collect and use data for disaster prediction and relief.

II. Nature of data provided

OpenDRI does not offer raw data through its platform, but rather three types of collaborative projects which involve the dispensation and use of environmental data. These are:

i) Sharing data projects with the goal of providing support to governments for creating locally owned open data platforms. There are 32 projects in this category listed, spanning across Central and South America, Europe, the Middle East, Africa and Asia.

ii) Collaborating with local communities and governments for the use of mapping tools to create accurate and updated data about their cities. There are 16 projects spanning Central America, Africa and Asia.

iii) As part of a collaborative project OpenDRI has worked on developing InaSAFE which is a free software which can be used to simulate disaster scenarios. There have been nine projects spanning across Africa and Asia.

Furthermore the platform includes the sharing of information through guides and publication of reports and training materials which are freely available.

III. Sample project

OpenDRI was part of a collaborative project managed by the Tanzanian Commission for Science and Technology. Humanitarian aid organisations, the local community and university students worked together to map the flood-prone areas in Tanzania. They used OpenStreetMap to create maps of the region's floodplains, roads, streams, residential areas and more. These maps will help the local authorities have better disaster response during flooding events, as accurate maps of this type did not exist beforehand. Furthermore, this project helped raise awareness and trained local residents to use mapping tools which are skills they can then apply to other projects as well.

The maps produced by the project are freely available online, as well as in printed form for the local authorities. They provide information about the geography of the region as well as features relevant for relief, such as the materials the buildings are made of. Due to the open access availability of the data collected, it has been able to be reused for other purposes such as the mapping of the local transit system of the city.

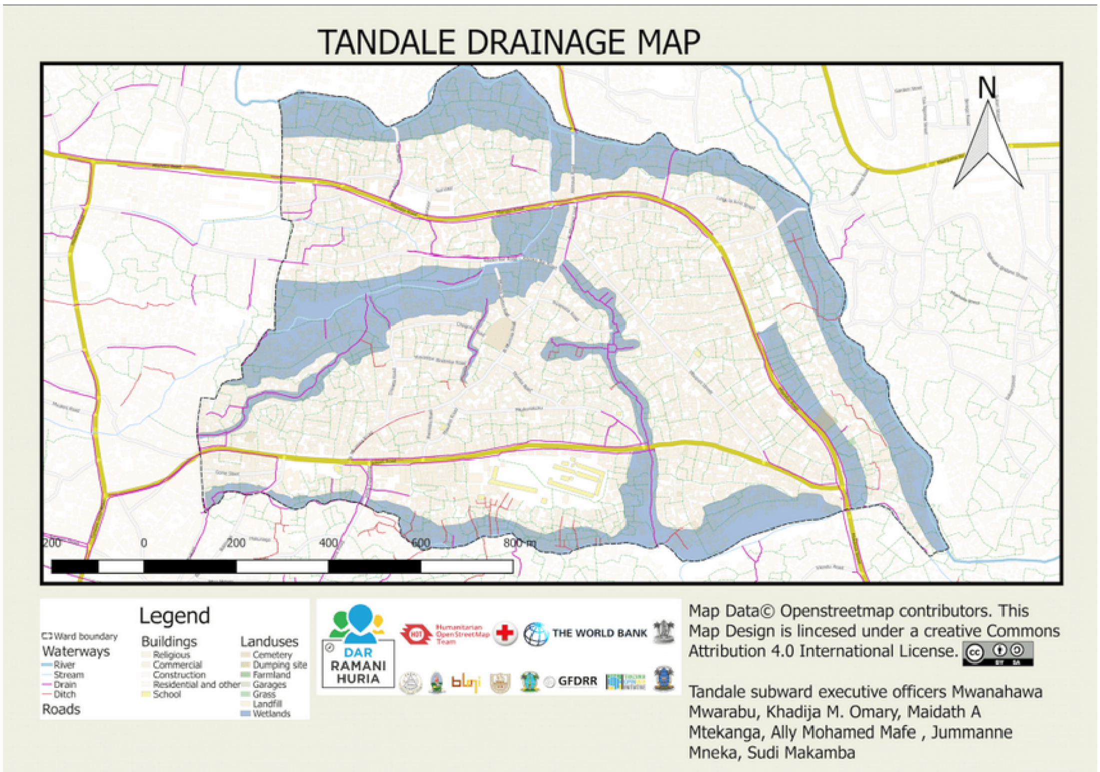


Fig 8. Note. From *Participatory Mapping for Historical Flood Inundation Extents.* 2016. (https://opendri.org/participatory-mapping-for-historical-flood-inundation-extents/)

4.2.2 Digital Green

Available at: https://www.digitalgreen.org/

I. Introduction to Digital Green

Founded in 2006 as part of a project being conducted in Microsoft Research. The project aims to work directly with farmers by providing training and information for more sustainable farming. FarmStack, a tool developed by Digital Green, provides farmers directly with up to date relevant data. Information about the weather, soil and seeds, as well as recommendations, are dispersed through channels such as messaging services, videos and chatbots. Data is therefore being directly delivered to the relevant parties so as to make real time decision making more efficient on a local level.

II. Nature of data provided

Digital Green’s data products are highly targeted and specialised to the project. There is no raw data made available on their platform, but the projects showcased include soil, weather and seed data which is updated and delivered to farmers in various ways depending on the technologies available and needs of the individuals involved.

III. Sample project

In a project based in India focusing on cashew farmers, Digital Green worked on providing weather information to farmers to help prevent flower drop and flower burn. Weather conditions like drought and fog, as well as certain funguses, regularly contribute to the loss of cashew plants in the region. Through the use of highly localised weather information and soil analysis, the network disperses information through video and voice messages, to account for a low literacy rate in the region.

4.2.3 Earth Observation for Sustainable Development (EO4SD)

Available at: https://www.eo4sd-drr.eu/

I. Introduction to EO4SD

This project is managed by the European Space Agency (ESA) in conjunction with international financial institutions (IFIs) such as the World Bank, European

Investment Bank, Inter-American Development Bank, International Fund for Agricultural Development and the Asian Development Bank. It is aimed at providing Earth Observation data and actionable analysis to the development projects run by these institutions. The main focuses are Urban development, Agriculture and rural development and Water resources management.

II. Nature of data provided

Geographically, this project focuses on four regions in the world where disasters are more frequent. These are Africa, South-Asia, East-Asia Pacific and Latin America. Data is not made directly accessible through their platform, but rather they work directly with the partners and client countries.

III. Sample project

For World Bank's project City Resilience, wherein they invest in and promote cities developing their resilience to natural disasters, EO4SD provided geographic data and analysis for specific locations under the project named "WB-City Resilience Program: Terrain deformation in urban areas." Nine cities in total were assessed through this project, namely Banjul, Barishal, Beira, Cap Haitien, Georgetown, Khulna, Paramaribo, Vinh Long and Yangon. Through this project geographical data was able to be quickly prepared and dispersed so that actionable insights for city development could be shared with the respective authorities, making it possible for these cities to increase their resilience to extreme events.

These projects provide a blueprint for the potential that environmental data technology has on the outcomes of people in the real world. Focusing on specific geographic regions and working closely with the local communities allows the optimisation of the use of this information to solve the most pressing problems. On the other hand, these large projects are time consuming, temporary and are not always going to be available for every region or climate problem.

4.3 Data made available directly

The following data sources provide data directly through their platforms, either in raw or analysed form, or both.

Furthermore, the first group presented here will involve the NGO sector, meaning organisations, partnerships or groups that do not work directly for countries' governments or for profit within the industry. Within this group are Water Data Portal, GARDIAN and Digital Earth Africa.

Within the group of governmental organisations, the World Bank's data project will be presented, alongside the United Nations' (UN) World Environment Situation Room, the United States' Data portal, and the European Union's European Environment Agency and Copernicus Programme.

For the private sector, Microsoft's AI for Earth will be showcased.

This thesis will attempt to take into consideration how these different sources of funding and access to technology might affect the way the platforms are organised and how the data is presented, as well as the diversity and goals of the available projects.

4.3.1 Water Data Portal (WDP)

Available at: http://waterdata.iwmi.org/

I. Introduction to WDP

This is the platform for the provision of data under the organisation International Water Management Institute, an NGO working in water research based in Sri Lanka. The goal of this organisation is to serve as a hub for mapping services and provide a mix of technical tools and products alongside research and analysis.

II. Nature of data provided

WDP provides direct access to data relating to water and agriculture on its website. The variety and features of this data has been provided in greater detail in Table 1 below:

Table 1. Overview of data products, WDP

Name	Type	Details	Geography
Global Irrigated Area Mapping	Interactive map, Maps	Mapping global irrigated and rainfed cropland using satellite images	Worldwide
Irrigated Area Map Asia and Africa	Interactive map	Mapping irrigated cropland in data poor regions	Asia, Africa
River Basin	Map	Creation of baseline layers for researchers	Worldwide
Eco-Hydrological Databases	Table	Various data extracted from journals, reports, unpublished documents, experts' opinions and internet sites	Select countries
Global environment flow information system	Table	Estimating the volume of water needed to maintain freshwater-dependent dent ecosystems	Worldwide
Global Drought Patterns	Interactive map	Global patterns and indicators of drought	Worldwide
Glacier and snow in Asia basin	Maps, Tables, Graphs	Assessment of the storage properties of glaciers	Asia
Flood Risk Mapping	Interactive map	Mapping the vulnerability to flooding of different regions in Asia and Africa	Multiple countries
Nepal Climate Change Vulnerability	Interactive map	Identifying water basins in Nepal and assessing their vulnerability to climate change	Nepal
Water Quality Mapping - Jaffna	Interactive map	Assessing the chemical properties of the Chunnakam aquifer	Sri Lanka

Agro-Well Mapping Jaffna	Interactive map	Identification and georeferencing of agro-wells as farmers return to the region after war	Sri Lanka
Rainwater harvest calculator	Interactive calculator	Determining the needed size for rainwater storage tanks	Worldwide

There are a total of 25 products and tools presented on their website, this selection showcases the main categories so as to avoid repetition of similar projects. Some of the provided tools have broken links and images, showing they have not been recently updated. Appendix B shows an example of this.

III. Sample project

In the interactive map shown below in Figure 9, a visualisation of the agricultural wells in the area of Jaffna, Sri Lanka is presented. The tool works well and showcases a lot of relevant information, with options for personalisation and exploration of different components of the datasets. These maps are part of the Agro-Well Mapping Jaffna product, included in the table above.

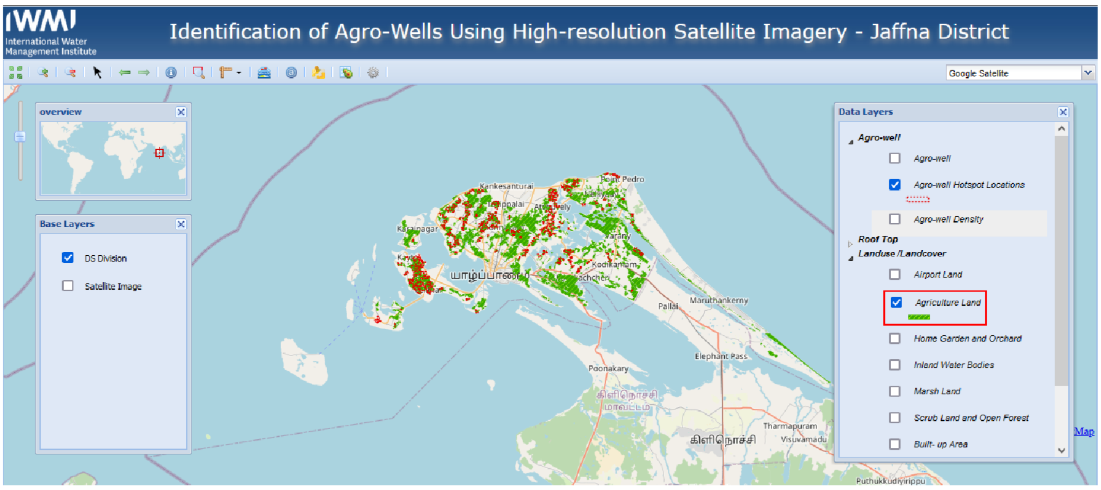


Fig 9. Agricultural land layered alongside well locations. Generated by author. http://waterdata.iwmi.org/Applications/Jaffna_AgroWell/ (20.03.2022)

Here there is an overlay of maps with the placement of agricultural land and wells. This project showcases the advantages of having a highly localised and specialised data project. The information is detailed and diverse, and therefore highly applicable for particular problems that could be tackled in the region.

The user interface of the platform is easy to navigate. However, there are some issues with the presented tools, mostly in the form of a lack of platform maintenance. Many tools have not been updated regularly, which leads to the issue of certain maps not working properly as images don't load. In some cases, the maps work on one browser (Mozilla Firefox) but not on others, such as is the case with the Global Irrigated Area Mapping tool. Furthermore, for some of the datasets, the raw data is quite old and therefore potentially not very useful for newer projects, considering that certain things might change over time such as the development of urban areas. Furthermore, some datasets in the form of tables were not presenting data, indicating either incomplete datasets or mistakes in the upload.

However some of the tools work quite well and are easy to use. For example, the Rainwater Harvest calculator is easy to adjust and information is provided explaining the formulas used for the tool.

4.3.2 Global Agricultural Research Data Innovation & Acceleration Network (GARDIAN)

Available at: https://gardian.bigdata.cgiar.org/

I. Introduction to GARDIAN

The Consultative Group for International Agriculture (CGIAR), is an international conglomerate of organisations that specialise in food security. As part of this partnership, the GARDIAN platform was created in 2019 as a means to share resources through the entire data cycle, including the collection, curation as well as analysis of datasets. The data itself is gathered from multiple sources.

II. Nature of data provided

The platform provides worldwide data, the earliest of which is from 1970's and is still being updated this year in 2022. The types of data provided include geospatial and tabular, as well as that related to research in genetics.

Data is easy to sort through geographically through the interactive map which also informs the user about the size of the datasets for each region. Furthermore, it is an intuitive interface for filtering through publication year, type of data or keywords. Furthermore, the platform provides search analytics, meaning users can easily see additional components such as which country in the selected region has the most research papers, which are the key words and phrases, who are the biggest data providers and funders for the area, and what kind of projects are being done by CGIAR.

However there is not much room for the optimisation of the search, and considering the breadth of data including that of research papers and journals, sifting through it to find relevant data might be a challenge. The mapping tools, although helpful in showcasing a quick overview, are not highly customizable. Furthermore, a lot of the data is only available off platform in its raw form, as GARDIAN has a higher focus on serving as a collector which aggregates various data projects in one place to be searched, rather than direct data products ready for use.

III. Sample project

Below, Figure 10 shows the mapping tool with the yellow layer indicating the harvested area, as seen in 2017.

Fig 10. Harvested Area (2017). Generated by author. https://gardian.bigdata.cgiar.org/#/search/ (19.03.2022)

Furthermore, below in Figure 11, the page that GARDIAN shows for each dataset on its platform is shown. Each dataset's authors and geographic focuses are showcased, as well as an evaluation of how much the data follows FAIR indicators (assessing whether the data is Findable, Accessible, Interoperable and Reusable) as evaluated using the Netherlands Institute for Permanent Access to Digital Research Resources (DANS) metrics for FAIR compliance, a set of standardised guidelines for the collecting and sharing of data.

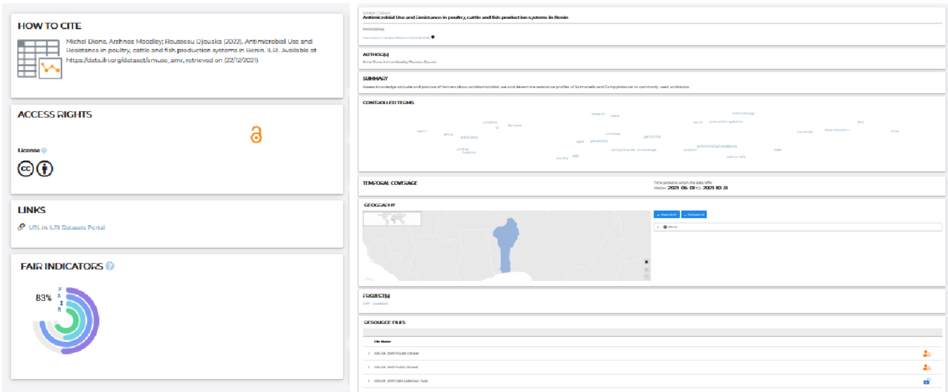


Fig 11. GARDIAN's presentation of individual datasets. Image generated by author (https://gardian.bigdata.cgiar.org/#/asset/04b4158bf6b31989f3b7b86471d261f3) (19.03.2022)

4.3.3 Digital Earth Africa

Available at: https://www.digitalearthafrica.org/

I. Introduction to Digital Earth Africa

A platform which makes use of spatial data to provide insights and support to governments and individuals on the African continent specifically. They provide both raw data which can be used by others as well as analysis ready data.

II. Nature of data provided

There are a range of tools and resources which are made available through this platform, including raw spatial data, mapping tools for visualisations, code for analysis so that it can be replicated, as well as analysed and visualised data with particular focus on different aspects of spatial analysis, such as water or crop observations, and composite images. These tools are easy to use and up to date.

The topics addressed specifically are diverse in nature. They are Satellite Data for Sustainable Development, Climate Action and Reporting, Water Resources and Flood Risks, Agriculture and Food Security, Land Degradation and Coastal Erosion and Urbanisation. Each of these categories is presented in detail, its importance in the African continent and which of their tools are meant to tackle which of these problems.

The data they provide is based on Landsat and Sentinel satellite imagery which goes through machine learning algorithms which measure reflectance and NDVIs (Normalised Difference Vegetation Index, a measure of vegetation density). They present this data alongside support for deeper understanding of both the potential and the limitations of it in the shape of training, guides, a helpdesk and an online course to their tool available in both English and French.

III. Sample project

A large data catalogue is easily accessible through their platform. The design of the platform is intuitive and easy to navigate. They have a variety of tools and support available, including an open Slack channel for asking questions directly. Additionally, they present additional projects which they worked on for specific impacts.

One example is the visualising tool Digital Earth Africa Map. Data can be uploaded as well as searched within their database. The interface is easy to use and the data is divided into clear categories Satellite images, Surface water, Agriculture, Vegetation, Land cover, Meteorology and Elevation. Figure 12 below shows the tool's visualisation of the monthly rainfall averages. Additionally, the metadata of the dataset is clearly presented and easy to check.

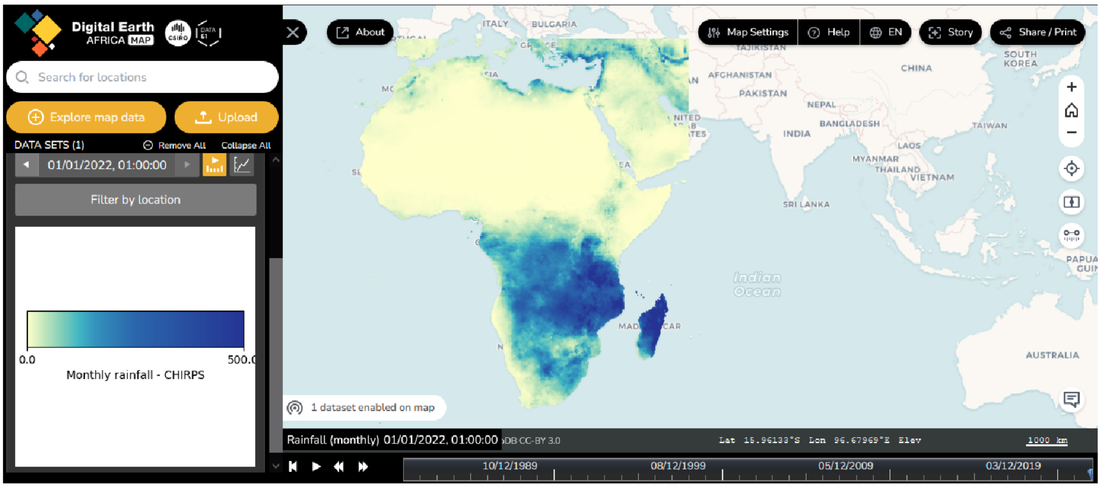


Fig 12. Digital Earth Africa Map, monthly rainfall. Generated by author. https://maps.digitalearth.africa. (16.03.2022)

Furthermore, a collaboration project with the Tanzanian National Bureau of Statistics to combat and prepare for extreme weather events is presented. Tanzania is vulnerable to droughts and floods so it is important for authorities to develop an understanding of the water systems in place.

For this project, mapping of Lake Sulunga was done due to its vulnerability to changes in water level by tracking its changing boundaries through time. Figure 13 below shows the tracking of the size and shape of the lake in the year 2020. The lake is of high importance to the local communities as it provides water for both drinking and agriculture practices. By collecting data about the area they were able to provide insights which helped locals know where to plant crops or to move away from certain areas as they were deemed more vulnerable to flooding or to be less fertile for agriculture.

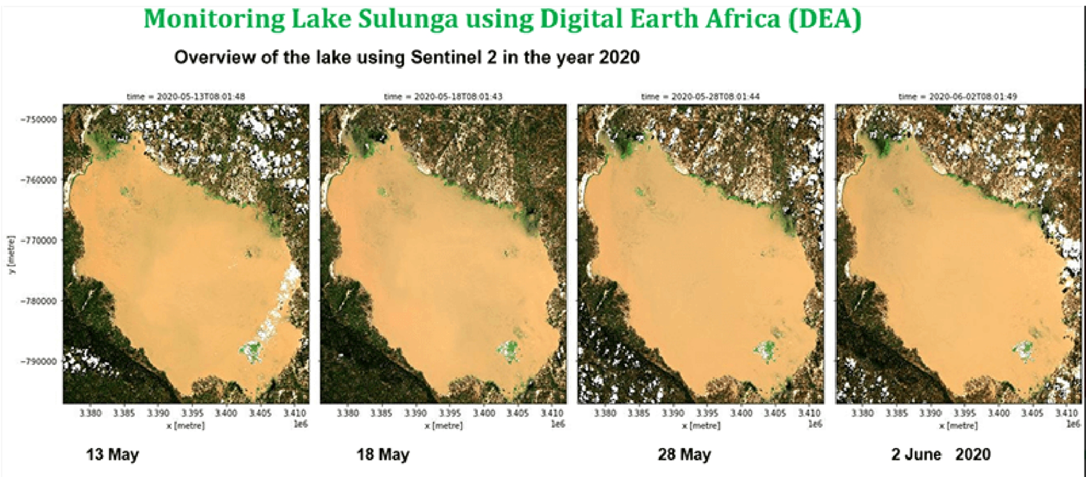


Fig. 13 Note. From *Using satellite data to combat drought: Monitoring Lake Sulunga, Tanzania.* 2020. (https://www.digitalearthafrica.org/why-digital-earth-africa/impact-stories/using-satellite-data-combat-drought-monitoring-lake-sulunga)

4.3.4 Data Bank

Available at: https://data.worldbank.org/

I. Introduction to Data Bank

Working as a collection spot for datasets from multiple international institutions, the data made available by the World Bank is easy to sift through. The main collaborators for this are the International Bank for Reconstruction and Development (IBRD), the International Development Association (IDA), the International Finance Corporation (IFC), the Multilateral Investment Guarantee Agency (MIGA), and the International Centre for Settlement of Investment disputes (ICSI).

II. Nature of data provided

The data is not specialised in environmental science, but rather serves as an aggregate for datasets across different topics where data is collected such as social indicators, the financial sector and more. There are 20 total indicators provided by the Data Bank of which four are directly related to environmental science. These are Agriculture and Rural Development, Climate Change, Energy & Mining and Environment.

There are 31 datasets under Environment, 35 under Climate Change, 15 under Energy and Mining and 21 under Agriculture. Besides graphs of time series data, there are also maps providing geospatial information on a country basis. All data is available for download for individual use.

The data are majority time series, though not always up to date with majority datasets ending in between 2015-2018. However the data are visually presented with the choice to view the data as different types of graphs, as well as split into countries. Metadata is also provided, as is the raw data for download. The Data Bank visualisation tool gives additional functionalities, such as deciding which countries to include in the aggregation, as well as change the styles and layout of the charts, and other simple functionalities which are helpful for creating personalised visualisations.

III. Sample project

Within the Climate Change category, data on a per country basis release of greenhouse gases is available. Graphing on country level is also made easily accessible, showcased in Appendix C.

The base graph is shown in Figure 14 below:

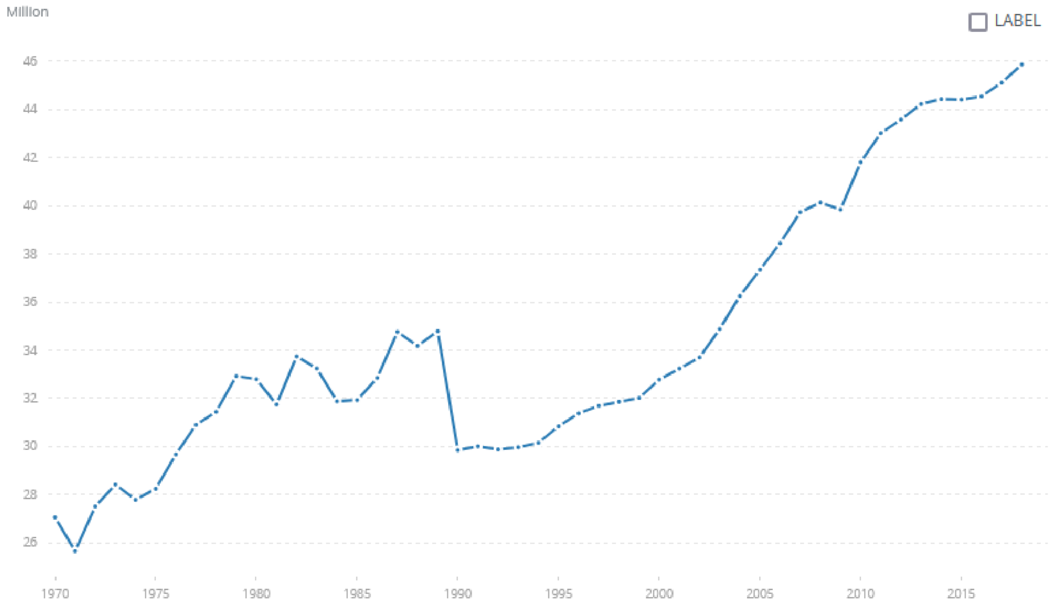


Fig 14. Note. From *Total greenhouse gas emissions (kt of CO2 equivalent).* (https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE?view=chart) (21.03.2022)

Meanwhile, the below graph Figure 15 shows the alternative way to showcase the same data, which is in the form of a choropleth map:

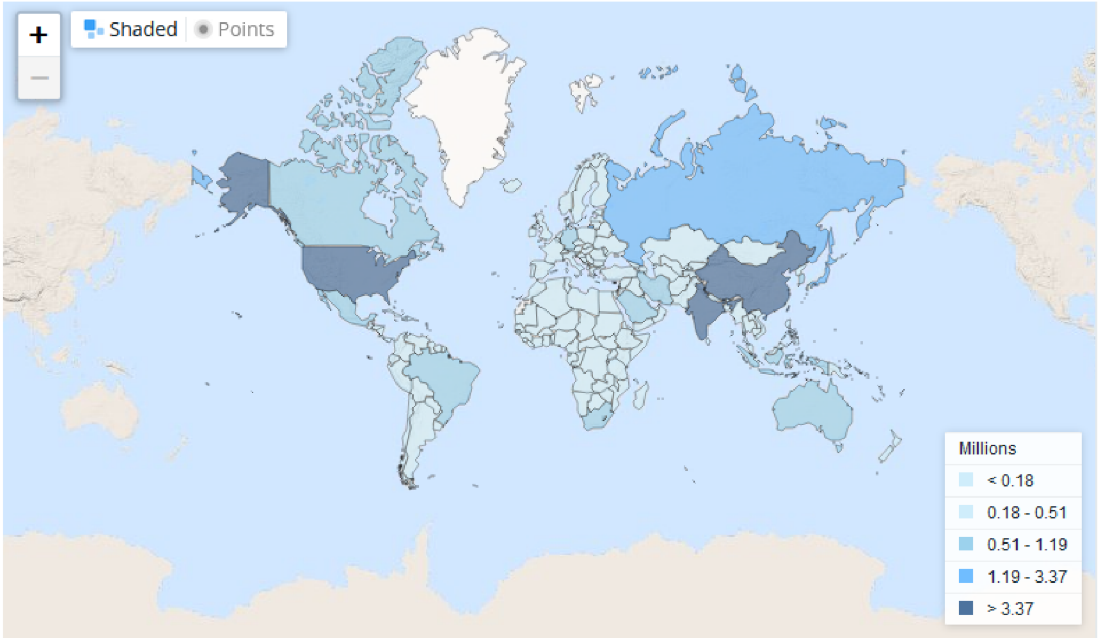


Fig 15. Note. From *Total greenhouse gas emissions (kt of CO2 equivalent).* (https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE?view=map) (21.03.2022)

The Data Bank therefore provides its data in a well presented way and information on a country level which is useful for global comparisons, alongside some customization options which are quite valuable when looking for specific information for more localised projects. However there is not a wide range of data topics covered and some datasets have not been updated in several years.

4.3.5 World Environment Situation Room

Available at: https://wesr.unep.org/

I. Introduction to the World Environment Situation Room

A branch of the UN founded in 1970 and specialising in gathering, collecting and presenting environmental data. This is a platform where visualisations and analysis of various variables which describe the state of the climate are provided, such as temperature or CO_2 emissions, as well as geospatial data which can easily be accessed and visualised through their mapping service.

II. Nature of data provided

The datasets are geographically spread out worldwide and for the most part, temporally up to date to this year, 2022. The interface is easy to navigate, the data highly diverse and immediately available for use through their visualisation tools.

Topics include Biodiversity and Nature Loss, Climate Change, Pollution and Waste, Air, Chemicals and Waste, Indigenous Knowledge, Mountains, Oceans, Seas and Coasts, Ozone Layer Protection, Resource Efficiency, Risk, Sustainable Consumption and Production, Transport, Water, Biosafety, Disasters and Conflicts, Ecosystems, Energy, Environmental Rights and Governance, Environment under Review, Extractives, Forests, Gender, Green Economy and Technology. Not all topics have datasets within them at this time.

Within Climate Change, data are divided into a geospatial map and time series of various variables (CO2 concentrations, sea level rise, etc). Each of these is visualised through graphs on a global level and the raw data is available. Although there is some variation between datasets, most charts showcase data up to 2020 or 2021.

The geospatial mapping tool is easy to use and quickly visualises on a country level to showcase variables like temperature, sea level rise, wildfire frequency, extreme weather, and more. The Biodiversity section provides tools for gaining insights on the state of species counts and protected areas across the globe. The Pollution section has a map which visualises the levels of contamination across the globe, divided into types of pollution as well as specific pollutants and chemicals. The Air section showcases in real time air pollution across the globe, including World Health Organisation (WHO) recommended levels as well as categories of severity and impact on health from "GOOD" to "HAZARDOUS". The Oceans, Seas and Coasts section provides data on sea level rise, pollution as well as a map depicting chlorophyll levels across the globe. Ozone layer protection provides data on the historic and current state of the ozone layer as well as a per country report detailing the release of the chemicals which are known to be hazardous for the depletion of the ozone layer. The Water section provides links to mapping tools for the visualisation of existing water systems in the world as well as an in-site dashboard for a quick assessment of water quality and the levels of chemicals from 1962 to 2020.

III. Sample Project

The graphs below showcase the specific data mentioned above under the Climate

Change section, namely the temperature variable time series data in Figure 16, and Figure 17 the flood risk areas through a geospatial visualisation.

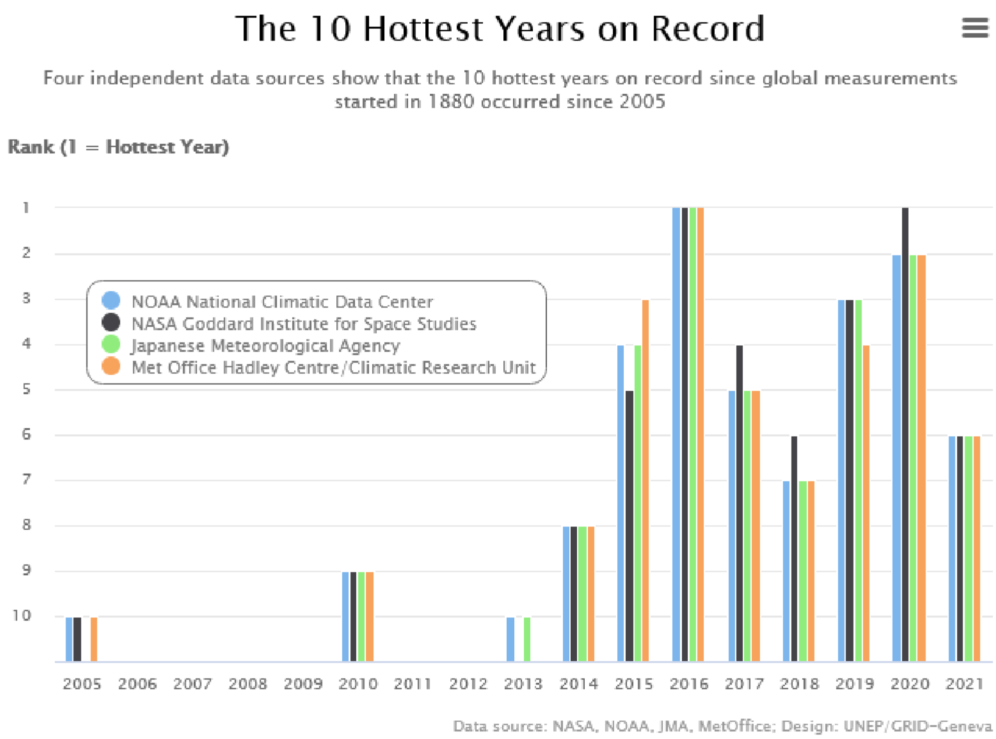


Fig. 16. Note. From *Global Temperature Change.*
(https://wesr-climate.unepgrid.ch/essential-climate-variables-ecv/global-temperature-change) (24.03.2022)

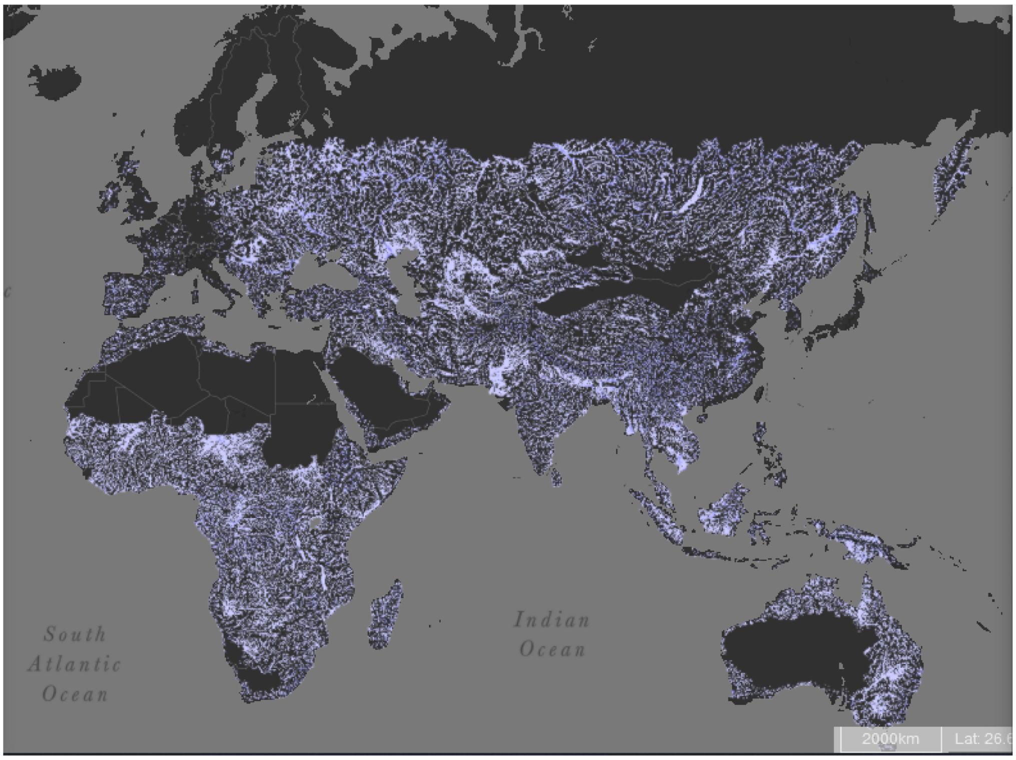


Fig 17. Note. Flood Hazard (25m). From *WESR: Climate.* Generated by author.
(https://app.mapx.org/?project=MX-5Z8-45E-K4I-SKH-75H&language=en) (21.03.2022)

4.3.6 European Environment Agency

Available at: https://www.eea.europa.eu/data-and-maps

I. Introduction to European Environment Agency

An agency of the European Union tasked with monitoring and providing data regarding the environment and aid in sustainable development.

II. Nature of data provided

The data, both of geospatial and time series type, is easily searchable and accessible, as well as presented in a way that would help users immediately gather insights through visualisations. The categories which are available and the number of datasets per category are detailed in Table 2 below:

Table 2. EEA data categories

Dataset Category	No. of datasets available
Agriculture	2
Air Pollution	8
Biodiversity - Ecosystems	11
Climate change adaptation	12
Climate change mitigation	16
Energy	6
Environment and health	4
Industry	6
Land use	3
Resource efficiency and waste	3
Soil	1
Sustainability transitions	2
Transport	7
Water and marine environment	9
Total	90

The data is strictly within the borders of the EU, which means that information for areas outside this jurisdiction is not available.

III. Project sample

Within the Water and marine environment category, a study of flood prone areas within the European Union given two climate scenarios is presented below in Figure 18, while Figure 19 compares the return period of sea levels between two climate scenarios:

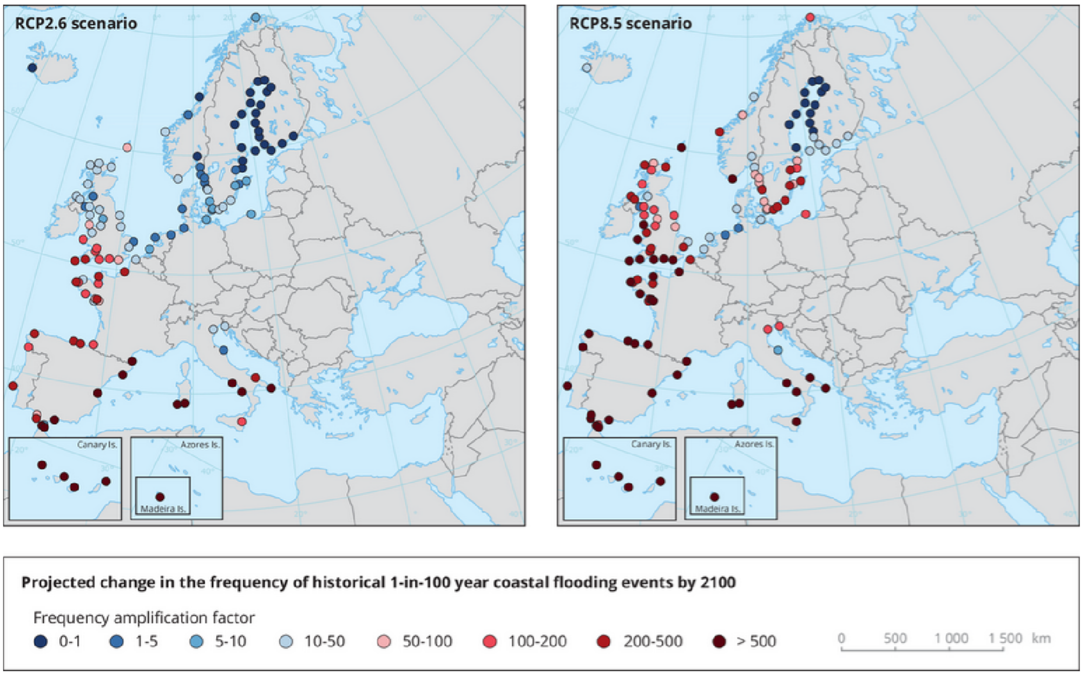


Fig 18. Change in the frequency of flooding events in Europe given projected sea level rise under two climate scenarios. Note. From *Extreme sea levels and coastal flooding.* (2021). (https://www.eea.europa.eu/ims/extreme-sea-levels-and-coastal-flooding)

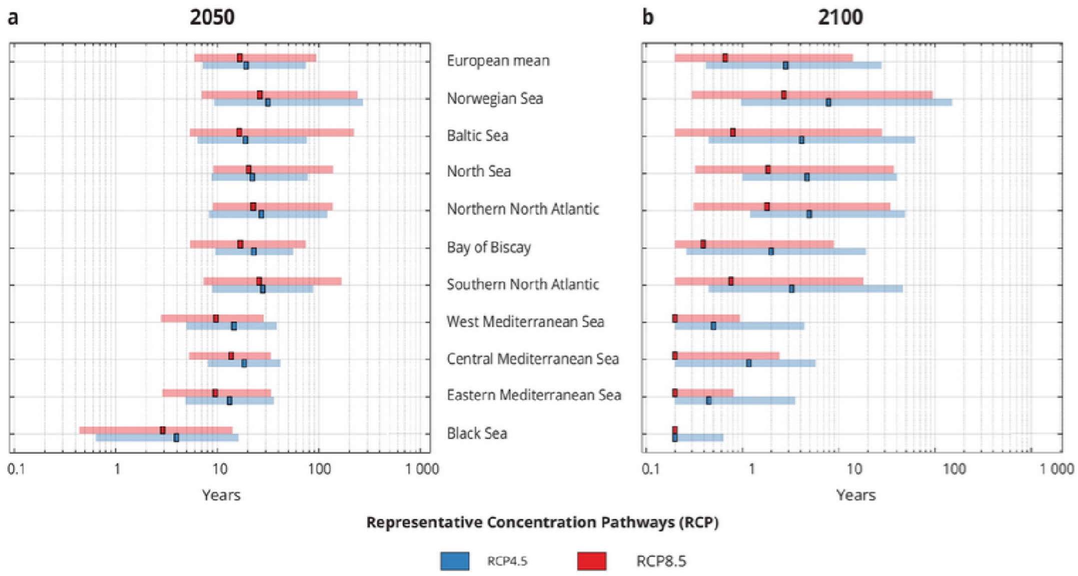


Fig 19. Return period of current 100-year extreme sea levels under two emissions scenarios. Note. From *Extreme sea levels and coastal flooding*. (2021). (https://www.eea.europa.eu/ims/extreme-sea-levels-and-coastal-flooding)

Background information on the issue of flooding due to rising sea levels is elaborated on, alongside the maps showcasing the hotspots which are at risk for coastal flooding. The raw data is referenced, alongside the metadata and references which are made available.

4.3.7 The Copernicus Programme

Available at: https://www.copernicus.eu/en

I. Introduction to Copernicus

The European Union's Earth Observation program, managed by the European Commission. The observations collected include both satellite as well as in-situ data, and are made freely available to the public. Copernicus is currently the largest provider of space data in the world, with an average of 16 terabytes of data being delivered on a daily basis.

II. Nature of data provided

The Copernicus program has several services which it makes available for use freely to users online. Table 3 below provides an overview of the main services which are related to natural disasters and climatological extremes specifically:

Table 3. Copernicus Program services overview

Name	Details
Copernicus EMS (Emergency Management Service) On Demand Mapping: Rapid Mapping	Geospatial information is provided within hours or days after a natural disaster to provide support in the disaster management process
Copernicus EMS On Demand Mapping: Risk and Recovery Mapping	Geospatial information for the support of disaster management, including assessing vulnerability and prevention of various natural hazards
Copernicus EMS Early Warning and Monitoring: European Flood Awareness System (EFAS)	Monitoring and prevention of extreme precipitation events on a global scale. Mapping services provide information about flood risks, prevention as well as emergency response. Complementary flood forecasts are also provided by their service Global Flood Awareness System (GloFAS)
Copernicus EMS Early Warning and Monitoring: European Forest Fire Information System (EFFIS)	Geographically focused on Europe, the Middle East and Africa, EFFIS monitors wildfire events in real time, providing support for wildlife management
Copernicus EMS Early Warning and Monitoring: The Drought Observatory (DO)	The Drought Observatory works on a European and Global scale (EDO and GDO respectively) and specialises in providing reports in response to expected drought events
Climate Data Store	Works as an aggregation system where climate data can easily be searched and filtered. Raw data, metadata and documentation are all provided in an intuitive and easy to search way. Data is also made available through an API, and includes instructions on how to make an API request.
Climate Data Store Toolbox	A programming interface to find, analyse and display data. Code samples, documentation, application use cases are provided making it easy for users to familiarise themselves with the platform and its possibilities.

This overview shows that the Copernicus Program has a versatile set of services and data products relating to the environment in general, and natural hazards and climatological extremes in particular, which it makes available to decision makers, researchers and interested parties alike.

III. Sample project

Developing a catalogue of high impact weather events in conjunction with historical records of loss and damages through the Pluvial Flood Risk Assessment for Urban Areas Project. Through this project, they are combining different types of time series, historical and geospatial data to provide insights to individuals and decision makers regarding the risks associated with extreme precipitation events which lead to flooding without the presence of an overflowing water body. These insights are helpful for city planning and disaster preparedness.

Figure 20 below shows the workflow used to create the flood risk analysis with the use case of the city of Vienna, Austria.

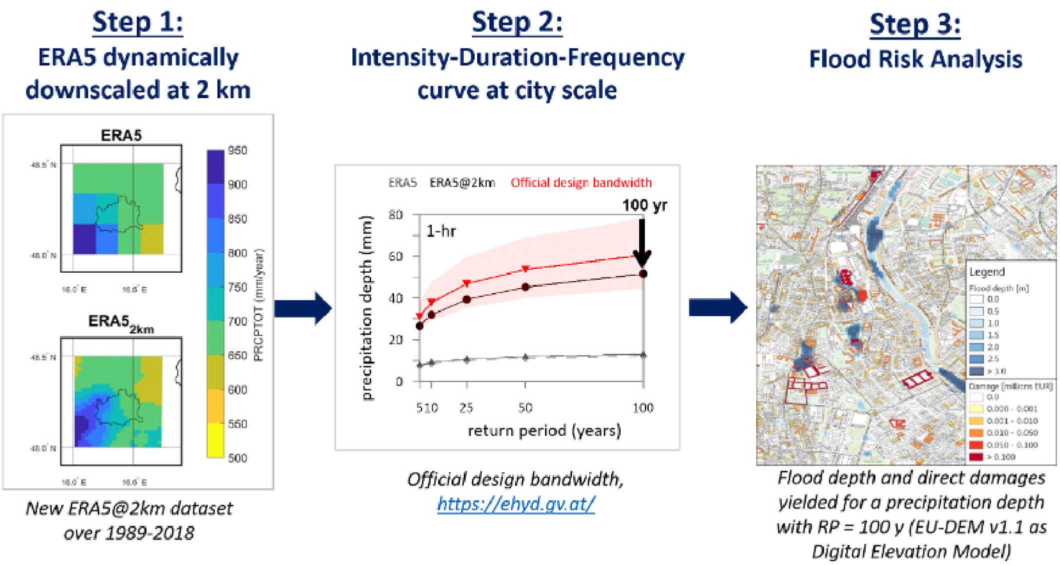


Fig 20. Example of the modelling chain set for the flood risk analysis: city of Vienna. Note. *Pluvial Flood Risk Assessment in Urban Areas*. (https://climate.copernicus.eu/pluvial-flood-risk-assessment-urban-areas)

The Copernicus Program in this way stands out as a large scale data provider, while also working on specific projects based on geography or type of natural hazard being addressed, as well as developing knowledge sharing through their platforms such as in the case of open source code and examples, and reporting and analysis being made freely available.

4.3.8 Data.gov

Available at: https://data.gov/

I. Introduction to Data.gov

Created and managed by the government of the United States of America, since 2009 the online resource data.gov has provided a conglomerate of datasets from various government sources which operate within the country, including agencies, cities, states and universities. It is meant as a one stop shop for data collected by these agencies, as they are required by law to provide their data openly to the public.

II. Nature of data provided

Data is up to date and covers the territory of the United States, as well data collected about other countries. The search engine is easy to navigate through different categories, date ranges, locations or agencies as providers of data. There are also detailed instructions provided for users in the Help section of the website.

There are a total of 343,360 datasets found at the time of this writing (20.03.2022). The largest single category of data is Local Government with 16,687 datasets, which are localised data collected by the government on its region. Under the Topic of Climate Change there are 409 datasets, of which 259 are of geospatial type. 72 are under the Energy tag, nine under Ocean, seven under Maritime and four under Agriculture.

However the way the data search is organised is not optimal. Tags for the datasets do not provide a big overview of related words but focus on the most specific ones. For example, searching for “Africa” does not result in every search that has to do with Africa but only those datasets that have Africa in the title. As such it is difficult to browse this data catalogue without knowing in a highly specific way what the user is searching for.

III. Project sample

As a data aggregation system, this source has fewer visualisation options than the other platforms presented earlier. However it is helpful for finding particular information when it is available through these governmental sources. The figure below is a flood risk map taken from the Global Flood Hazard Frequency and Distribution, created by the Dartmouth Flood Observatory.

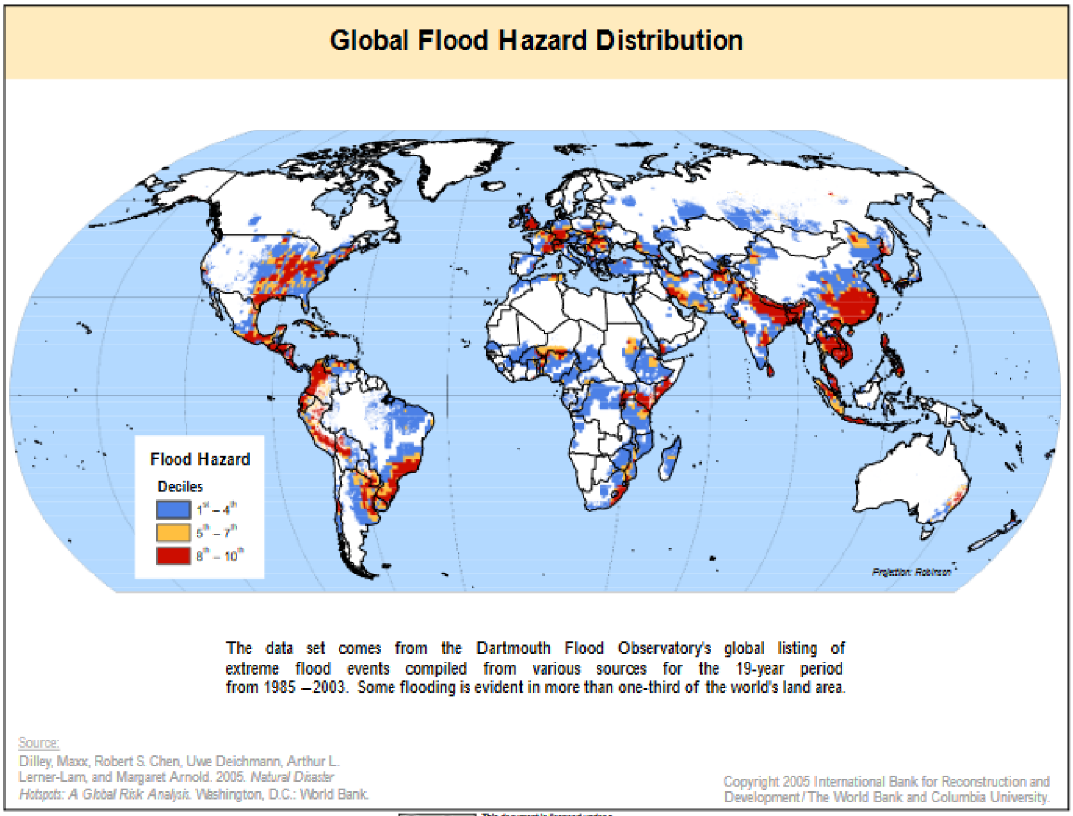


Fig 21. Global Flood Hazard Frequency and Distribution (Center for H, Risk Research CCU and Center for International Earth Science Information Network CCU, 2005)

Information provided by the platform includes not only access links to the original research and maps, but the metadata, references and copyright information.

4.3.9 AI for Earth

Available at: https://www.microsoft.com/en-us/ai/ai-for-earth

I. Introduction to AI for Earth

A program which began as a Microsoft project and utilises their artificial intelligence (AI) technology, Azure, focusing on solving environmental problems using data. The data is aggregated and made available through their data catalogue, named Planetary Computer.

They have four projects showcased which are focused on combining data and Azure technology. These are SOS Mata Atlantica with a focus on water quality in Brazil, Ocean Data Platform which tracks pollution from ships, Imazon which is focused on deforestation of the Amazon rainforest, and lastly ThermaFY which monitors changes in temperature.

II. Nature of data provided

Their data catalogue offers prepared datasets from satellite images, there are a total of 31 currently. Additionally to the metadata, the code used to process the images is also provided in the form of examples, making it possible for users to learn and apply the same techniques

III. Project sample

In the Imazon project, satellite imagery is used to detect risk factors such as unofficial roads. By sifting through the satellite images using Azure Machine Learning technology, large amounts of data can be analysed in a short amount of time, allowing for quicker response and better preparation in the wake of deforestation. The ability to quickly and accurately assess data in this way has potential for disaster relief and response.

5. Discussion

The groups of data sources provided some variety in the way they approached collecting and working with data. A more detailed breakdown is provided in the sections below:

5.1 Providing data through agreements and providing data openly on the web

The first two selected data sources, OpenDRI and Digital Green, worked directly with users without making the raw data openly accessible on their platforms. The third, EO4SD, worked with client governments and agencies through collaboration with the institutions which run development projects in the region. These organisations had highly specialised projects which worked on solving specific problems in certain regions. Part of their work also includes the data collection process itself, which is not the case in all the other open data platforms presented. This type of data sharing project provides benefits through a personalised approach, and the integration of local communities in the solving of problems which affect them the most.

Furthermore, by involving the local community the data was able to be interpreted more meaningfully. The information provided was actionable for local decision makers, creating a clear line between providing data and affecting outcomes, such as the case with weather data which was used to aid farmers in protecting their crops in India. In the projects which involved educational courses for the use of specialised tools, the students were able to gather knowledge which they could then go on to apply to other projects as well. In the case of Digital Green there is the additional component of developing technologies for the dispersal of the data to the people who need it the most, while taking into consideration their individual needs. EO4SD made it possible to optimise the use of satellite technology for projects which were already in the making, increasing the potential for those institutions and governments to make decisions with a better understanding of their particular resilience development needs.

On the other hand, the other sources provided data directly on their platform, or served as an aggregator where data could be found. These provide access to data to a wider range of people, increasing the possibility of informing others in a meaningful way. However the information itself can be difficult to find, depending on the platform and how they are optimised. Oftentimes, the data is not current, and metadata is not always present. Sometimes the links are broken, showcasing a lack of website maintenance, which highlights the importance for upkeep in this industry. Furthermore, the data itself is more generalised, meaning the insights provided are not always going to be as deeply varied and meaningful for a particular area as they would be when a lot of data is collected with a focus on a particular place or problem.

5.1.1 NGO sector

Data platforms from the non-governmental sector have varied data and cover a wide scope of geographical regions. Water Data Portal and Digital Earth Africa had narrower scopes of geographical regions covered than the others. Digital Earth Africa's focus on one continent provided a combination of highly specialised data alongside a wide range of possibilities and data products which are openly accessible to users. Additionally they invested in teaching tools, which contributes to the spread of knowledge in the region.

Water Data Portal on the other hand had some projects which were from various parts of the world or global, but many were focused on Sri Lanka. However the platform suffered from poor upkeep and old data. GARDIAN works more as an aggregator of data, however their filtering options were intuitive and therefore easy to use, alongside some limited visualisation tools.

5.1.2 Governmental sector

The European Environmental Agency had a similar structure to Digital Earth Africa, in that they provided their data with a geographic focus on the borders of the European Union. Their data is presented rather in the form of reports, than as platform tools that allow for personalisation and research. This has the advantage of making it easy to find interpretations and analysis of the data as well, which can in some cases be more meaningful than the raw data or visualisations alone. However without many options for explorations the breadth of data available is more narrow compared to other sources. Furthermore, the Copernicus Programme has a much larger scope in both geography, as it works with data and organisations from regions outside the European Union as well, and data products such as raw data, analysis and platform solutions are all made available. This makes it possible for the vast resources made available through this programme to be used in varied ways to solve different kinds of problems for a greater number of people.

Similarly to GARDIAN, Data.gov provided a search engine for data gathered by many different agencies and organisations. However it was difficult to navigate and use of the data required some foreknowledge as it wasn't directly accessible and ready for use on the platform itself. The World Bank's Data Bank and the UN's World Environment Situation Room have additional benefits due to being the platforms of well funded and established organisations that operate worldwide. Their data was diverse and easily accessible, available for insights and customisation. Data Bank's platform also included data from other industries, which would make it comparatively easier to combine with climate data in search of new insights, since they are presented in a compatible way on a single platform.

5.1.3 Industry

Within the private sector, Microsoft's project AI for Earth is making use of environmental data while implementing machine learning technology developed at the company. The use of this specialised technology which is not made freely available outside their platforms opens the door to tackle additional challenges, such as the deforestation monitoring project provided in the example.

On the other hand, the data and visualisations they provide are not very diverse or high in quantity, but the source code of the workflow is made available which is a valuable teaching tool for interested users.

5.2 Needs of researchers

With regards to the needs the researchers presented in the survey shown in the Literature Review section, the data has shown varying degrees of compliance. GARDIAN, Data.gov and the Climate Data Store, functioning as specialised search engines for data which include published research, fill the need for researchers to be able to sift through large amounts of journals to find specific data, given that the survey showed journals were the main way they looked for relevant data. The downside here is that the data aggregated are coming from specific partners only, whether from the network of NGOs as is the case with GARDIAN, or including only research done by governmental bodies as is the case with Data.gov.

The platforms consistently provided metadata, which the survey results highlighted as one of the researchers' top needs. The data on the platforms was for the most part easy to discover and visualise for insights, as well as available at no cost.

In terms of the needs of researchers, these sources provide many of the most important factors they listed as important. While a more detailed aggregation system might be helpful, generally the discoverability of data through these platforms works well. The biggest issue is that some of the platforms are inconsistent with their upkeep and updates of the datasets.

5.3 Natural hazards, climatological extremes

The nature of the data varied among the sources, but the most common natural hazard which was possible to look up within most of the platforms was flooding and droughts. Climatological extremes in the form of precipitation and temperature were the most common datasets consistently available. Flooding hazards were presented in multiple platforms, as well as in projects focusing specifically on protecting land and communities from these extremes.

5.4 Modelling

In terms of modelling, the platforms at times had several options such as the case with EEA's visualisation of two different climate change scenarios, or when the possibility to showcase and compare data from several studies is made possible for visualising such as in the case of GARDIAN's datasets for crops, among others. The platforms were consistent in providing references to the research which was being used in their visualisations and interpretations., allowing users to check the assumptions, data and models being applied and compare when needed.

The datasets were often available in the same formats, either as CSV or XLSX files for tabular data, and most commonly GeoTiff and ESRI for geospatial data. This allows for a degree of interoperability which was highlighted as an important component of open access data in the literature review section.

6. Conclusion

The platforms presented here were diverse and had individual strengths and weaknesses provided by the technology they use, available funding as well as the focus on specific problems they were addressing.

This project has shown that there is a high degree of diversity in the way that environmental data can be applied and presented. Some platforms focused on providing the data itself, or information about the data alone. Others on presentation and interpretation or analysis, while others on participating in projects tackling specific problems, as well as in developing technologies for the dispersal and optimal use of the data and incurred insights. A couple of the platforms were also involved in the data collection process while others, rather than collecting it individually, focused on aggregation itself.

Additionally, several of the platforms engaged in knowledge sharing both in terms of analysing the data which was presented, as well as source code and help navigating their tools.

While no one approach is better than the others objectively, Digital Earth Africa and the Copernicus Programme had the strongest balance between diversity of data, presentation and teaching tools, alongside participating in the tackling of environmental problems through collaborative projects. The private industry showcased through Microsoft's AI for Earth was an example of how advancing data science technology can be used to tackle ecological problems in particular. Governmental agencies presented data in clear and easy to navigate ways alongside explanations and interpretations of data, which are highly useful for individuals looking to learn more about the environment.

This thesis aimed to present some examples of the many existing environmental data providers online. Although these platforms provide valuable insights through their data products, they still constitute a very small portion of the data landscape currently available which is additionally constantly expanding. While some of the platforms used similar raw data in the form of Landsat and Sentinel satellite imagery, the way they were presented was still varied, showcasing how much room there is for analysis and development of these tools in the environmental sector and in aiding informed decision making.

Something that could be helpful for researchers, decision makers and interested individuals in the future to make better use of these tools is a catalogue or dashboard of providers. This kind of product could potentially aid in sifting through the available material more efficiently, so that users are able to find the most relevant information quickly rather than sift through multiple platforms. This would be a difficult challenge to take on given the volume of data available, and further research is needed to assess whether it would be possible.

7. Bibliography

1. Avisse, N., Tilmant, A., Müller, M.F. and Zhang, H. (2017). Monitoring small reservoirs' storage with satellite remote sensing in inaccessible areas. *Hydrology and Earth System Sciences*, 21(12), pp.6445–6459.
2. Belcastro, L., Marozzo, F., Talia, D., Trunfio, P., Branda, F., Palpanas, T. and Imran, M. (2021). Using social media for sub-event detection during disasters. Journal of Big Data, 8(1).
3. Bernstein, J. and Kemp, K. (2020). The Role of Spatial Science in Environmental Case Studies: A Special Collection from the University of Southern California. *Case Studies in the Environment*, 4(1), pp.1–5.
4. Blair, G., Henrys, P., Leeson, A., Watkins, J., Eastoe, E., Jarvis, S. and Young, P., 2019. Data Science of the Natural Environment: A Research Roadmap. *Frontiers in Environmental Science*, 7.
5. Borowitz, M., 2017. *Open space : the global effort for open access to environmental satellite data*. Cambridge, MA: MIT Press.
6. Brbulescu, A. (2018). *Studies on time series applications in environmental sciences*. Cham: Springer
7. Brynjolfsson, E. and McElheran, K. (2016). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*, 106(5), pp.133–139.
8. Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J. and Waters, N., (2015). Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. *International Journal of Remote Sensing*, 37(1), pp.100-124.
9. Christidis, N. and Stott, P. (2021). The influence of anthropogenic climate change on wet and dry summers in Europe. *Science Bulletin*, 66(8), pp.813-823.
10. Dietze, M., (2017). Ecological Forecasting. New Jersey: Princeton University Press.
11. Duan, H., Zhang, G., Wang, S. and Fan, Y. (2019). Robust climate change research: a review on multi-model analysis. *Environmental Research Letters*, 14(3), p.033001.

12. Easterday, K., Paulson, T., DasMohapatra, P., Alagona, P., Feirer, S. and Kelly, M. (2018). From the Field to the Cloud: A Review of Three Approaches to Sharing Historical Data From Field Stations Using Principles From Data Science. *Frontiers in Environmental Science*, 6.
13. Gad-ElHak, M., (2009). Large-scale disasters. Cambridge: Cambridge University Press.
14. Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A. and Chakraborty, B. (2018). A review on application of data mining techniques to combat natural disasters. *Ain Shams Engineering Journal*, 9(3), pp.365–378.
15. Gouveia, N. (2000). Time series analysis of air pollution and mortality: effects by cause, age and socioeconomic status. *Journal of Epidemiology & Community Health*, 54(10), pp.750–755.
16. Hadi, Krasovskii, A., Maus, V., Yowargana, P., Pietsch, S. and Rautiainen, M. (2018). Monitoring Deforestation in Rainforests Using Satellite Data: A Pilot Study from Kalimantan, Indonesia. *Forests*, 9(7), p.389.
17. Harpold, A. (2015). Use of Lidar in Environmental Science. *Oxford Bibliographies Online Datasets*.
18. Hart, J.K. and Martinez, K. (2015). Toward an environmental Internet of Things. *Earth and Space Science*, 2(5), pp.194–200.
19. Holzworth, D., Huth, N. and de Voil, P. (2010). Simplifying environmental model reuse. *Environmental Modelling & Software*, 25(2), pp.269-275.
20. IPCC, (2012) – Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (Eds.) Available from Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 8RU ENGLAND, 582 pp. Available from June 2012
21. Jain, P.K., Quamer, W. and Pamula, R. (2018). Electricity Consumption Forecasting Using Time Series Analysis. *Communications in Computer and Information Science*, [online] pp.327–335. Available at: https://link.springer.com/chapter/10.1007/978-981-13-1813-9_33.
22. Knüsel, B. and Baumberger, C. (2020). Understanding climate phenomena with data-driven models. *Studies in History and Philosophy of Science Part A*, 84, pp.46–56.

23. Lahoz, W., Khattatov, B., Menard, R. and Springerlink (Online Service (2010). Data Assimilation : Making Sense of Observations. Berlin, Heidelberg: Springer Berlin Heidelberg.
24. Maliva, R. (2021). Historical Evidence for Anthropogenic Climate Change and Climate Modeling Basics. *Springer Hydrogeology*, pp.47–70.
25. Maspo, N.-A., Bin Harun, A.N., Goto, M., Cheros, F., Haron, N.A. and Mohd Nawi, M.N. (2020). Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review. *IOP Conference Series: Earth and Environmental Science*, 479, p.012038.
26. Mayer-Schönberger, V. and Cukier, K., (2013). *Big data*. London: Murray.
27. Morgan, J.L., Gergel, S.E. and Coops, N.C. (2010). Aerial Photography: A Rapidly Evolving Tool for Ecological Management. *BioScience*, 60(1), pp.47–59.
28. "Participatory Mapping for Historical Flood Inundation" (2016). OpenDRI. (https://opendri.org/participatory-mapping-for-historical-flood-inundation-extents/)
29. Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L. and Vanrolleghem, P.A. (2007). Uncertainty in the environmental modelling process – A framework and guidance. *Environmental Modelling & Software*, 22(11), pp.1543–1556.
30. Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K. and Wang, X. (2021). Deep Learning-Based Weather Prediction: A Survey. *Big Data Research*, 23, p.100178.
31. Restrepo-Estrada, C., de Andrade, S., Abe, N., Fava, M., Mendiondo, E. and de Albuquerque, J., (2018). Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Computers & Geosciences*, 111, pp.148-158.
32. Ritchie H. and Roser M. (2014, November). *Natural Disasters*. Our World In Data. https://ourworldindata.org/natural-disasters#citation
33. Roche, D.G., Granados, M., Austin, C.C., Wilson, S., Mitchell, G.M., Smith, P.A., Cooke, S.J. and Bennett, J.R. (2020). Open government data and environmental science: a federal Canadian perspective. *FACETS*, 5(1), pp.942–962.
34. Said, N., Ahmad, K., Riegler, M., Pogorelov, K., Hassan, L., Ahmad, N. and Conci, N. (2019). Natural disasters detection in social media and satellite

imagery: a survey. *Multimedia Tools and Applications*, 78(22), pp.31267–31302.

35. Schmidt, B., Gemeinholzer, B. and Treloar, A. (2016). Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. *PLOS ONE*, 11(1), e0146695.
36. Sebestyén, V., Czvetkó, T. and Abonyi, J. (2021). The Applicability of Big Data in Climate Change Research: The Importance of System of Systems Thinking. *Frontiers in Environmental Science*, 9.
37. Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L., Hegerl, G., Seneviratne, S., Vautard, R., Zhang, X. and Zwiers, F., (2017). Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and Climate Extremes*, 18, pp.65-74.
38. Staudhammer, C.L., Escobedo, F.J. and Blood, A. (2018). Assessing methods for comparing species diversity from disparate data sources: the case of urban and peri-urban forests. *Ecosphere*, 9(10), p.e02450.
39. van der Meer, J., Beukema, J.J. and Dekker, R. (2013). Using stochastic population process models to predict the impact of climate change. *Journal of Sea Research*, 82, pp.117–121.
40. Weyant, J. (2017). Some Contributions of Integrated Assessment Models of Global Climate Change. *Review of Environmental Economics and Policy*, 11(1), pp.115–137.
41. Woźniak, G., Dyderski, M.K., Kompała‐Bąba, A., Jagodziński, A.M., Pasierbiński, A., Błońska, A., Bierza, W., Magurno, F. and Sierka, E. (2020). Use of remote sensing to track postindustrial vegetation development. *Land Degradation & Development*, 32(3), pp.1426–1439.
42. Zhang, H., Hagan, D.F.T., Dalagnol, R. and Liu, Y. (2021). Forest Canopy Changes in the Southern Amazon during the 2019 Fire Season Based on Passive Microwave and Optical Satellite Observations. *Remote Sensing*, 13(12), p.2238.
43. de Sherbinin, A., Bowser, A., Chuang, T., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., Popescu, I., Shonowo, A. and Sivakumar, K., 2021. The Critical Importance of Citizen Science Data. *Frontiers in Climate*, 3.

8. List of Figures

Fig 1. Classification of disaster severity (Gad-ElHak, 2009).

Fig 2. Number of deaths from disasters (Our World in Data: Number of deaths from disasters (online) [cit. p6], available at <https://ourworldindata.org/natural-disasters/>)

Fig 3. Economic damages from disasters (Our World in Data: Economic damages from disasters (online) [cit. p7], available at <https://ourworldindata.org/natural-disasters/>)

Fig 4. Which attributes are most important to open data (Schmidt et al: Open Data in Global Environmental Research. (online) [cit. p.11] available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695/>)

Fig 5. How important is open data in your community (Schmidt et al: Open Data in Global Environmental Research. (online) [cit. p.12] available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695/>)

Fig 6. How do you discover data (Schmidt et al: Open Data in Global Environmental Research. (online) [cit. p.13] available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695/>)

Fig 8. Tandale Drainage Map (OpenDRI: Tandale Drainage Map (online) [cit. p19) available at: <https://opendri.org/participatory-mapping-for-historical-flood-inundation-extents/>

Fig 9. Agricultural land layered alongside well locations (WDP: Agro-Well Mapping Jafna (online) [cit. P24] available at: <http://waterdata.iwmi.org/Applications/Jaffna_AgroWell>

Fig 10. Harvested Area (GARDIAN (online) [cit. P26] available at: <https://gardian.bigdata.cgiar.org/#/search/>

Fig 11. GARDIAN's presentation of individual datasets (GARDIAN (online) [cit. P27] available at: <https://gardian.bigdata.cgiar.org/#/asset/04b4158bf6b31989f3b7b86471d261f3>

Fig 12. Digital Earth Africa Map, monthly rainfall (Digital Earth Africa: Monthly rainfall (online) [cit. P28] available at: <https://maps.digitalearth.africa>

Fig 13. Overview of the lake using Sentinel 2 in the year 2020 (Digital Earth Africa: Using Satellite Data to Combat Drought: Monitoring Lake Sullunga, Tanzania (online) [cit. P29] available at: <https://www.digitalearthafrica.org/why-digital-earth-africa/impact-stories/using-satellite-data-combat-drought-monitoring-lake-sulunga/>

Fig 14. Total greenhouse gas emissions, line graph (DataBank: Total greenhouse emissions (kt of CO2 equivalent) (online) [cit. P31] available: <https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE?view=chart>

Fig 15. Total greenhouse gas emissions, choropleth map (DataBank: Total greenhouse emissions (kt of CO2 equivalent) (online) [cit. P31] available: <https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE?end=2018&start=2018&view=map>

Fig 16. The 10 Hottest Years on Record (WESR: Global Temperature Change (online) [cit.33] available at: <https://wesr-climate.unepgrid.ch/essential-climate-variables-ecv/global-temperature-change>

Fig 17. Flood Hazard (25cm) (WESR: Climate (online) [cit. p34], available at: <https://app.mapx.org/?project=MX-5Z8-45E-K4I-SKH-75H&language=en>

Fig 18. Change in the frequency of flooding events in Europe given projected sea level rise under two climate scenarios (EEA: Extreme Sea Levels and Coastal Flooding (online) [cit. p36] available at: <https://www.eea.europa.eu/ims/extreme-sea-levels-and-coastal-flooding>

Fig 19. Return period of current 100-year extreme sea levels under two emissions scenarios (EEA: Extreme Sea Levels and Coastal Flooding (online) [cit. p36] available at: <https://www.eea.europa.eu/ims/extreme-sea-levels-and-coastal-flooding>

Fig 20. Example of the modelling chain set for the flood risk analysis: city of Vienna (Copernicus: Pluvial Flood Assessment in Urban Areas (online) [cit. P39] available at: <https://climate.copernicus.eu/pluvial-flood-risk-assessment-urban-areas>

Fig 21. Global Flood Hazard Frequency and Distribution (Data.gov: GLobal Flood Hazard Frequency and Distribution (online) [cit. P.41] available at: <https://catalog.data.gov/dataset/global-flood-hazard-frequency-and-distribution>

APPENDICES

APPENDIX A

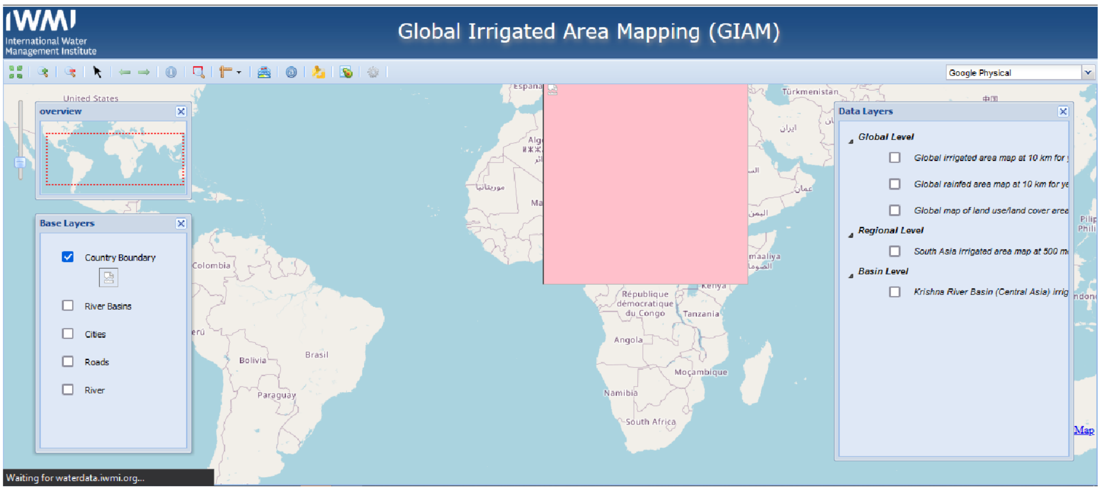
Raw data for Figure 7. Number of relevant research papers with keywords

	ecology AND (modelling OR modeling)	natural disaster AND (modelling OR modeling)	climate change AND (modelling OR modeling)
1980-1990	33,900	767	5,270
1990-2000	253,000	3,910	81,600
2000-2010	1,330,000	16,400	979,000
2010-2020	1,290,000	32,300	1,610,000

APPENDIX B

Screenshots from the data product presented in Table 1

Global Irrigated Area Mapping, broken link
(available at: <http://waterdata.iwmi.org/Applications/GIAM2000/giam.php>)



APPENDIX C

Per country visualisations by Data Bank make it easy to break down and quickly understand global data (available at: <https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE?view=chart>).

The screenshot below shows an example of this (using the same data as in Figure 14).

All Countries and Economies

Country	Most Recent Year	Most Recent Value
Afghanistan	2018	98,920
Albania	2018	10,080
Algeria	2018	218,910
American Samoa	1989	21
Andorra	2018	590
Angola	2018	79,730
Antigua and Barbuda	2018	1,210
Argentina	2018	365,650

Fig 22. Per country breakdown and visualisation of total greenhouse gas emissions