



**DATA 4
NATURE**



NATURAL
CAPITAL
IRELAND

Derryclare Lough, Co. Galway



Natural Capital Ireland hosted the first national 'Data 4 Nature' Workshop on May 11th, 2021. This event was funded under the Open Data Engagement Fund of the Department of Public Expenditure and Reform with additional funding from the Office of Public Works.



**An Roinn Caiteachais
Phoiblí agus Athchóirithe**
Department of Public
Expenditure and Reform



Report Contents

4 Introduction

- What is Natural Capital?
- Hosting a Data 4 Nature workshop

8 The current national and international nature-data context

15 Summary Recommendations

22 The workshop

- Workshop aims
- Who attended

28 Break-Out Groups, Challenges and Recommendations

54 Appendices

- Breakout room prompts
- Further information on data dashboards
- Relevant links & further reading



Introduction



**DATA4
NATURE**

Introduction to the host of Data 4 Nature

Natural Capital Ireland (NCI)'s mission is to promote the natural capital approach and working with stakeholders and partners is key to that mission. As part of this, NCI is a partner on the pioneering INCASE (Irish Natural Capital for Sustainable Environments) project which is trialling Natural Capital Accounting (NCA) in Ireland for the first time at catchment scale.

The INCASE project requires a broad swathe of nature-related data, as well as data on land cover and land use, to create a set of accounts that can tell a story about the condition of our natural capital and the ecosystems services it provides us, using the global standard for NCA: the UN System of Environmental and Economic Accounting, also known as the SEEA. Hence NCI is in a good position to have a broad overview of data issues, needs and gaps.



What is Natural Capital?

Natural Capital is an economic metaphor for nature – the stocks of nature (biodiversity, water, atmosphere and geosphere) that provide the flows of goods and services that benefit humans. Accounting for these helps us to measure the vital services that flow from nature to our society and economy, benefits such as the provision of food, climate regulation, landscapes for our recreational use and ultimately health, wellbeing and inspiration.

The natural capital approach has a strong policy backing in Europe. In late 2019, the European Green Deal was announced, which states that “all EU policies should contribute to preserving and restoring Europe’s natural capital”. The EU’s new Biodiversity Strategy states that by 2050, the EU’s “natural capital will be protected, valued and appropriately restored”.



Hosting a Data 4 Nature Workshop

Given the challenges of finding and utilising nature data on the INCASE project, Natural Capital Ireland applied to the Open Data Engagement Fund for support to host a national workshop on this topic.

The Data 4 Nature Workshop, held on May 11th, 2021, brought together policymakers, state agencies, data holders and academics to discuss ways in which the collection and publication of environmental, land and water data can be harmonised to facilitate the preparation of Ireland's first natural capital accounts.

In doing so, we were able to facilitate conversations and record the concerns and suggestions of those creating and utilising nature-data in Ireland. Based on those conversations, we prepared recommendations for improvements on how the data could be gathered, managed and presented - the recommendations are contained in this report.

The Current National and International Nature-Data Context



**DATA4
NATURE**

The Current Policy Landscape

Under the EU Open Data Directive public bodies are required to make data available for reuse, where possible, in open and machine-readable formats.

The upcoming UN Biodiversity Conference will have a range of cross-cutting issues on the agenda which will change how many look at nature-related data, including linking climate change & biodiversity.

The UN has formally adopted the SEEA-Ecosystem Accounting statistical standard for linking environmental information to national economic accounts which will revolutionise the way we calculate success in this area.

The Programme for Government has committed to a National Digital Strategy which will support Biodiversity Data Collection with plans for a strategic policy unit in each government department encompassing complex data analysis skills.

Ahead of COP26, the Climate Action and Low Carbon Development (Amendment) Bill will result in the introduction of sectoral carbon-budgets which will likely generate a lot of environment-related data from various sectors of society.

The Central Statistics Office has set up an ecosystem accounts division.



The Open Data Strategy & Open Data Directive

Open Data Strategy 2017-2022 aims to provide a framework for data “to be used to ensure that public services are delivered in a more effective and efficient manner”.*

The Open Data Directive, entitled ‘Open Data and the re-use of public sector information’, was transposed into Irish law as of July 16, 2021. Public bodies are required to make data available for reuse, where possible, in open and machine-readable formats.

The directive states that: “Dynamic data should be made available via APIs (Applications Programming Interfaces) and bulk download where relevant.” The Directive includes the concept of high value datasets (HVDs), which is defined as “data, the re-use of which is associated with important benefits for the society and economy”.

Certain lists of high value datasets have to be made available by public bodies and public undertakings for free in machine readable format, accessible via APIs and free to re-use. Publicly funded research data will now come under the Directive and new national policies will be required aimed at making this research openly available.**

The Open Data Unit recommends standardisation to an open format, with an open licence for reuse (CC BY 4.0).

*https://data.gov.ie/uploads/page_images/2018-03-07-114306.063816Final-Strategy-online-version1.pdf

**<https://data.gov.ie/blog/the-new-open-data-directive-what-will-it-mean-for-public-bodies>



The next few years are set to be a ground-breaking period for nature-related data:



The Open Data Strategy 2017-2022 will be reviewed in 2022 and we hope that this report will provide useful recommendations as the Department of Public Expenditure and Reform conducts research into the impact of the Open Data Strategy thus far.



A perception was expressed by workshop participants that nature-related data is currently not well used in the decision-making process – this keenly relates to the various challenges in utilising data, which we will examine.



Current state of nature-related data in Ireland

In Ireland there are a lot of high-quality datasets readily available, particularly noted was the Office of Public Works (OPW), the National Biodiversity Data Centre (NBDC), and the National Parks and Wildlife Service (NPWS).

- The National Biodiversity Data Centre often provides very detailed methodologies, vital for reproducibility. Biodiversity Maps were developed as a national mapping and data portal for biodiversity, and the Data Centre serves as Ireland's node to the Global Biodiversity Information Facility. It favours a national strategic approach to standardise, document and make biodiversity data accessible, according to FAIR Data Principles.
- The OPW data re-use and citation terms apply to make all data in its possession and ownership available, except in cases where there are specific reasons for not releasing data. This might include sensitivities around making environmental data freely available because the location of some protected species and habitats should be guarded.
- The datasheets which accompany maps from the National Parks & Wildlife Service were also highlighted as a critical component for the reproducibility of data.



Using the Open Data Portal for NCA

The INCASE project has made extensive use of the Open Data Portal (ODP), including data on catchments (EPA); soils, aquifers and bedrock (GSI and Teagasc); water quality (EPA); habitats and species (NPWS, EPA, Teagasc, Coillte, Forest Service, Local Authorities, NBDC, Heritage Council, etc.); CSO environment accounts and planning data (Local and Regional Authorities).

However, many other highly relevant environmental datasets are not currently available on the ODP, including data on land use (from Department of Agriculture and Teagasc), forest cover (Forest Service and Coillte), and peatlands (e.g., data from Bord na Móna).

There are currently 3,262 environmental datasets on the ODP (the second highest sector on the platform) and yet, as Ireland's environment faces severe crises on several fronts, efforts to harmonise these datasets to aid policy decision making have been lacking.



All data on natural systems in Ireland is collected for specific purposes such as EU reporting (e.g., Water Framework and Nature Directives), national reporting (e.g., forest statistics to FAO, GHG emissions under National Inventory Reporting), and/or relating to payments (e.g., the Land Parcel Identification System). This means that relevant datasets are gathered by different organisations, over different time periods, scales and for different purposes, using different methods to fulfil different criteria.

There is no central information point bringing together all the data on natural systems in Ireland. In order to be maintained into the future, our natural capital must be managed in a holistic and systemic way. We need data on all natural systems to properly understand the connections and trade-offs between them.

Our workshop found that there are many very good environmental datasets but with a large variety of datasets from different sources and data gatherers, users are faced with a set of key issues and our participants made some key recommendations on this.



Summary Recommendations

**DATA4
NATURE**

5 key recommendations

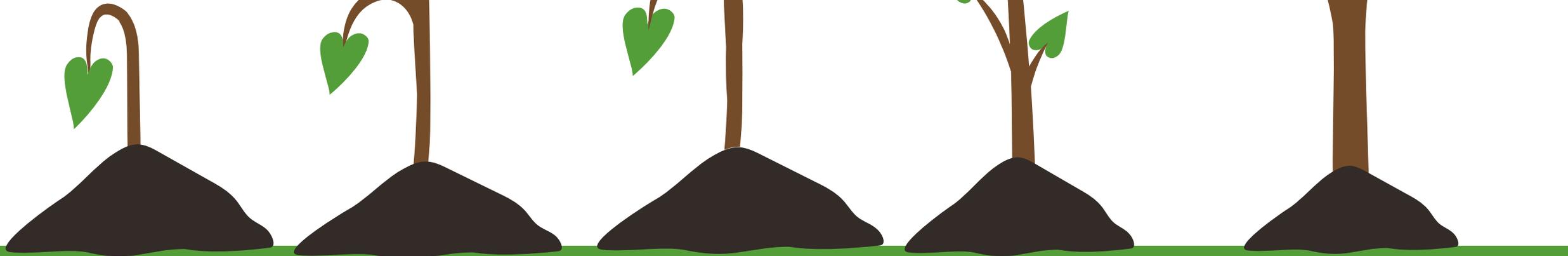
1. Develop a central dashboard for Irish nature data

2. Establish robust data gathering protocols

3. Harmonise formats

4. Publicise and promote

5. Ensure open & equal access



1. Develop a centralised dashboard for Irish nature data

- Present all of Ireland's nature data sources together within a single dashboard (utilising the [Open Data Portal](#) in coordination with existing sources) to enable the development of National Natural Capital Accounts. Inclusion of the forthcoming [OSI](#) land cover map will be key.
- Develop a common data management plan for state and semi-state bodies to establish principles and protocols for gathering and sharing of data to improve communication between state bodies. The plan should include the standardisation (where possible) of metrics, metadata, cleaning steps, coordinate systems, resolutions, minimum mapping units and time series for datasets and follow the [FAIR Data Principles](#). [The Open Data Unit](#) recommends standardisation to an open format, with an open licence for reuse ([CC BY 4.0](#)).
- Design data gathering steps using a 'multi-actor' approach to ensure engagement with all sectors involved in the process including landowners, farmers, citizen scientists, ecologists, data scientists, community groups, businesses, environmental NGOs and those working in the public sector.
- The dashboard should record and display the data protocols and methods used to collate and manage the data and provide very clear instructions on how to record data to ensure it is reliable and reproducible.
- Create a national environmental data organisation to verify all data uploaded to the dashboard.



2. Establish data gathering protocols for Irish nature data

- Metadata should be provided for individual Geographical Information System layers to allow for easy interpretation of field names.
- Cleaning steps should be scripted using code which can be published as part of the metadata. Ideally a package of raw data, cleaning code and standard data product (SDP), created using the cleaning code and raw data, would be produced.
- Correct formats should be identified for different kinds of data, excluding those such as PDFs which are unsuitable for open data management.
- An appropriate system should be developed for tagging data which will facilitate searches.
- Introduce procedures to ensure data quality where low-cost sensors are in use to guarantee the data produced with these technologies is interoperable with future datasets and complies with FAIR Data Principles.



3. Develop protocols to ensure harmonisation of formats

- Certain formats such as PDFs do not fall under the definition of 'Open Data' - the end-user and purposes of the data (primary and secondary uses) must be considered at the earliest stages when setting out to gather data. Major public bodies should provide a spatial framework for viewing datasets. Very clear metadata procedures, written in plain English and without jargon, following accepted standards, should be designed with non-specialist data users in mind.
- Establish coordination between the larger gatherers of environmental and nature-related data. Lack of coordination can result in lack of spatial and temporal alignment in data sets. For example, the INCASE project found that when using datasets from different agencies to map natural capital extent and condition of ecosystems in Ireland, they encountered several different coordinate systems, resolutions and minimum mapping units as well as different time series for datasets.
- Transferability of data is necessary. How formats may change in the future should be considered. Thought must be given to the importance of file standardisation, especially in complex databases, e.g., Excel, Access. SQL was mentioned as the standard approach for such databases now.
- The Open Data Unit recommends standardisation to an open format, with an open licence for reuse.



4. Establish methods to publicise and promote Irish nature data

- Invest in the promotion of the dashboard and the education of those who use it to establish buy-in from data-gatherers and users, including citizen scientists, government bodies and professional researchers. Support from politicians will be vital to establish and maintain a strong data strategy.
- Establish strict protocols on data privacy, security and governance and how best to engage people and organisations.
- Investment should be made in time, staff and funding to ensure innovative and holistic data practices are established and well-promoted to facilitate the long-term and future health and wealth of our data ecosystem.
- A central signposting system within the dashboard will promote ease of use and findability of data.
- Thought should be given to strong media and communications campaigns to raise awareness among the wider public on what data is available and how to access it.
- Appropriate support should be provided through the dashboard to help users with visualisation and data analysis.



5. Ensure equality of access

- All social groups must have equal access to data (taking into account the need for extra protections for sensitive data.) Standardisation to an open format, with an open licence for reuse ([CC BY 4.0](#)) is recommended by [The Open Data Unit](#).
- Open data guidelines must be adhered to by all stakeholders. This may require greater collaboration and coordination between large data centres and SMEs.
- Plain language should be employed in metadata, and for dashboard use instructions. The dashboard should demystify data by using common language.



The Workshop

**DATA4
NATURE**

The workshop took place online over one day and was divided into morning and afternoon sessions

Over 100 attendees

13 expert presentations

12 breakout-room discussions where all participants were invited to give their views on the potential for improving the availability, accuracy, usability, accessibility, reproducibility, and discoverability of nature-related data in Ireland

Notetakers relayed the main points raised in the discussions which were used to inform the recommendations in this report



Workshop Aims

1. To establish the state-of-play for nature-related data in Ireland
2. To facilitate a conversation between data collectors, publishers and users
3. To record the key themes and concerns of those working with nature data
4. To present recommendations to Minister of State with special responsibility for Public Procurement & e-Government Ossian Smyth based on workshop outcomes



Who attended?

Members of the Public

Major bodies responsible for Irish nature data

The National Parks and Wildlife Service (NPWS),
The National Biodiversity Data Centre (NBDC),
and The Office of Public Works (OPW)

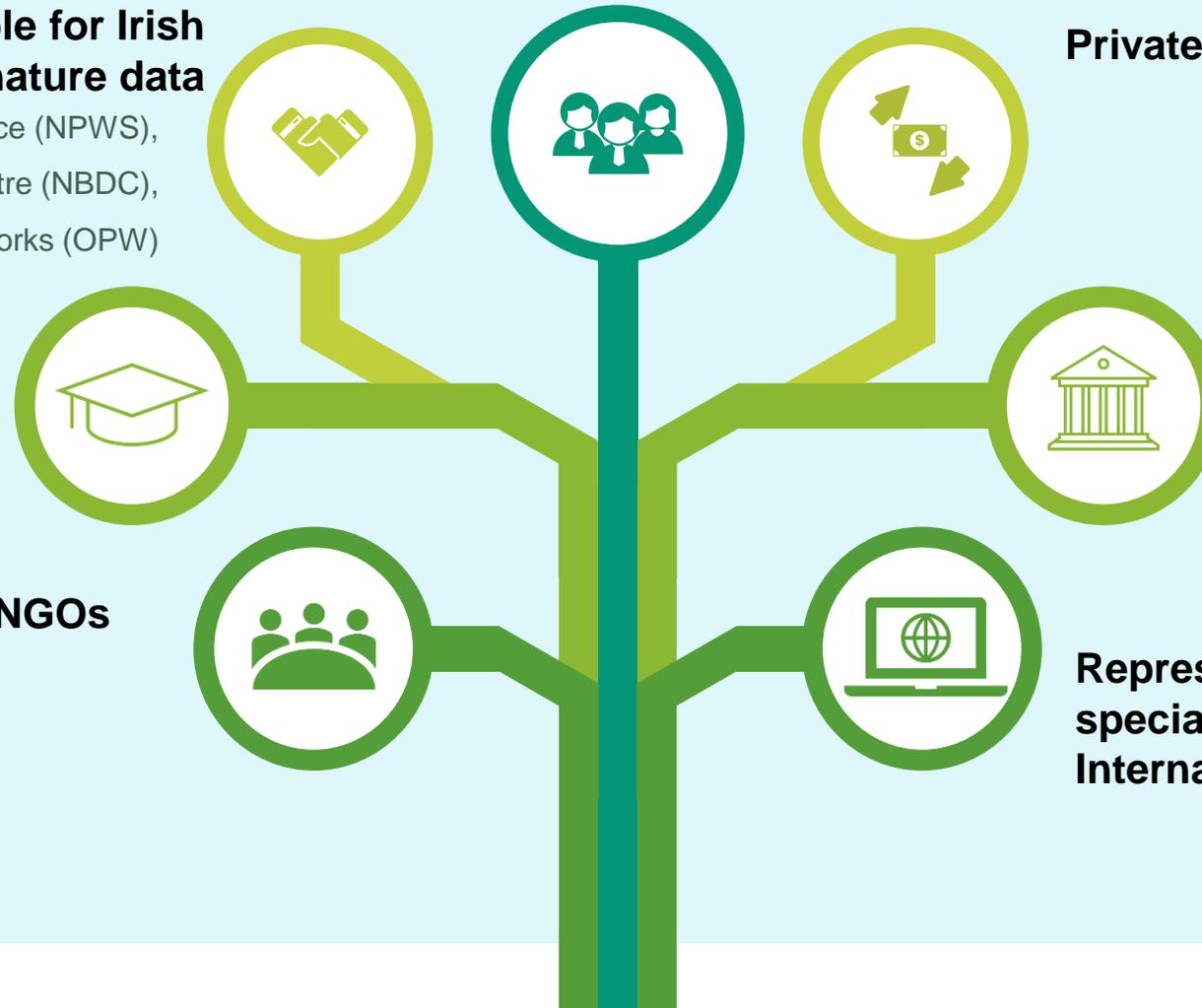
Private Sector

Academia

Public Sector

NGOs

Representatives of specialised Irish and International projects



Morning Session

- Dr Edward Curry, Co-Principal Investigator at the Insight Centre for Data Analytics
- Professor Yvonne Buckley of Trinity College Dublin and the All-Island Climate and Biodiversity Research Network
- Gemma Weir, NPWS
- Q&A

- Recordings of workshop sessions are available to watch here: <https://www.naturalcapitalireland.com/data4nature>



Afternoon Session

- Dr Liam Lysaght, Director of the National Biodiversity Data Centre
- Paddy Morris, EPA Catchments Unit
- Tony Brew, Environmental Section, OPW
- Lisa Coleman, INCASE
- Luke Binns, Dublinked
- Ainhoa Gonzalez, ESM Webtool
- Rowan Fealy & Tim McCarthy, Terrain AI
- Kevin Doolin, Demeter
- Breakout Room Discussions & notetaker reports
- Q&A
- Rapporteur Kieran Sullivan of AgriDISCRETE project



Break-Out Groups, Discussions and Recommendations

Attendees were assigned to small breakout groups to discuss their experiences and highlight issues they have encountered* with one of the following categories.** Analysis of these discussions yielded our recommendations detailed in this section.

Data availability

Data accuracy

Data usability

Data accessibility

Data reproducibility

Data discoverability

*Some attendee quotes have been recounted in the following slides

**Prompt questions used for each category can be found in Appendix 1

1. Availability

- With the rise of social media and apps, more and more data is becoming available to us in general, but our Data 4 Nature stakeholders pointed out that, often, being aware of the existence of some of this data in the first place depends on who you know - and accessing data can be dependent on having the right contacts.
- Good directories are those which are well-maintained and updated – some examples cited by our groups included the [EPA GIS Portal](#) * & [UCD Research Repository](#).
- New open databases are currently being compiled from social networking, citizen science uploads and crowdsourcing apps and some of these, such as [STRAVA](#) and [Flickr](#), can be used in the study of cultural services, which has useful implications for the natural capital approach as it can help us begin to measure hard-to-quantify aspects of natural systems such as recreation, leisure, scenic value, health and wellbeing.*

*See 'Assessing cultural ecosystem services through crowdsourced data from social media' Edited by Johannes Langemeyer, Bonnie Keeler, Derek Van Berkel, Andrea Ghermandi
<https://www.sciencedirect.com/journal/ecosystem-services/special-issue/10TJJXMV25L>



“Online searches are a confusing rabbit-hole of various links - we need easier-access formats to make data findable & accessible for the non-expert”

- **Availability Challenge 1:** Environmental data often comes via disparate datasets, with varying metrics and in hard-to-access formats - this can prove awkward for scientific researchers and data experts but is particularly difficult for the less specialist user.
- **Recommendation:** That a standardised set of principles to guide data-gathering be adopted. Agencies such as the CSO already use principles of data gathering as pre-defined by the UN, including adhering to data protection laws (GDPR) and safeguards in place to ensure confidentiality and privacy.*

*<https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/informationfordataproviders/>
https://unstats.un.org/unsd/dnss/gp/Implementation_Guidelines_FINAL_without_edit.pdf



“In order to promote nature tourism, leisure and conservation projects, for instance those relating to Natura 2000 sites and their benefits, we need data that is publicly available and understandable”

Availability Challenge 2:

A lack of awareness of what data is available in the first place.

Recommendation:

Better publicity and promotion of what nature-related data is available via a better data infrastructure, e.g. a data platform for nature that allows easy access to data from multiple areas and sectors in a much more aligned way.



A Note on Data Gaps

The lack of accurate data on land cover has prevented accurate monitoring and assessment of land use and therefore the ability to accurately predict and prevent pollution, flooding and other effects caused by climate change, urban sprawl and agricultural practices.

Data gaps, particularly when it comes to land-use mapping, were found to be a particular issue when it comes to putting together accurate natural capital accounts.

However, workshop participants raised the point that the forthcoming OSI land cover map will go a long way to filling in these gaps, with a much broader land classification system and more detailed resolution which will facilitate calculation of carbon emissions and sequestration and inform future land management.*

* <https://www.irishtimes.com/news/science/ireland-needs-needs-more-detailed-land-use-maps-1.4010070>



2. Accuracy

The quality of data in Ireland in general is deemed sufficient, however adequate coverage in terms of time and space is an issue for nature-related data. Sensors are used for both citizen science and institutional/ 'smart city' applications and, though they have many benefits, they also present a challenge in terms of data quality, validation and future interoperability of datasets.

- **Challenge 1:** Accuracy is an issue when it comes to citizen science, where non-experts and untrained members of the public are doing the bulk of the data collection, leaving more scope for error or incorrect collection methods, e.g. supplying incomplete metadata.
- **Recommendation:** For greater accuracy, care must be given at a fundamental level to ensure that the process of data collection and verification is set up rigorously. There is a need for a baseline on what data is available, where, and how it was gathered. Protocols must be in place to ensure the quantity, access, and overall accuracy of the data, while minimising error and accounting for bias.



Accuracy Challenge 2:

Low-cost sensors have an important role in citizen engagement activities as they empower non-experts with tools to make sense of their surrounding environment. Low-cost sensors for environmental and biodiversity monitoring are still in their infancy and present challenges of data quality, reliability, accuracy, and robustness, as well as future interoperability. (The need for a proper interpretation from experts must be considered prior to their use in citizen science as there is a lack of personnel with the taxonomic and ecological expertise to verify such data, particularly relating to identifying invertebrates, for example.)

The goals of ubiquitous sensing are a) to provide personalised information to citizens about their environmental impact, motivating engagement in sustainable practices; b) to capture data from a wide geographical area to improve spatial and temporal availability of data with respect to conventional sensors in field studies.

Recommendation:

It is critical to have procedures in place to ensure data quality where low-cost sensors are in use to guarantee the data produced with these technologies is interoperable with future datasets and complies with FAIR principles.



Accuracy Challenge 3:

- There is a lack of confidence in the accuracy of data relating to the quality and value of some of our ecosystems, which has implications for nature conservation and natural capital accounting. SMEs looking to invest in green measures to enhance certain areas have found that not enough relevant data is available on ecosystem extent and condition to inform such projects. This can lead to a lack of investment in these areas.
- While information may be available on the suitability of certain habitats for carbon sequestration, for example, there may be a lack of coordinated information on biodiversity and co-benefits such as water quality or public health. This can lead to uncertainty when it comes to carrying out conservation and biodiversity-enhancing actions. More focus must be put on collating this kind of information, including making use of social media innovations to examine cultural benefits.*

Recommendation:

- Relevant datasets should be aligned and accessed in a single dashboard geared toward national natural capital accounts, with lots of excellent data already available from National Biodiversity Data Centre, NPWS, CSO, OPW, Marine Data Centre and other sources.
- It would be helpful for this resource to record and display the data protocols and methods used to collate and manage the data to ensure it is reliable and reproducible.

* <https://besjournals.onlinelibrary.wiley.com/doi/full/10.1002/pan3.10199>



3. Accessibility

Data was recently dubbed ‘the world’s most valuable resource’ by the Economist*. There is no doubt that it is vital that all social groups have equal access to data. As it stands, often larger organisations have more access to information they can use to extrapolate insights to support their work and derive economic benefit.

A vital first step in accessing data is gaining knowledge of what datasets are available, next comes knowing where to access them and, most of all, having the capabilities to access and understand them. This applies in areas of research, education and citizen science. It was noted at our workshop presentation that connecting our data is going to be increasingly important as we become more dependent on new technology and Artificial Intelligence (AI) systems. We heard how the concept of a digital ‘ecosystem’ has come to the fore as we see the need to connect data systems together, to create ‘data value chains’ via these AI-powered systems.

*<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>



“One of the key challenges we have as a society is that data is not equally accessible by everyone”

Accessibility Challenge 1:

Datasets are not currently equally accessible by all sectors in many cases.

Recommendations:

A user-friendly open-data dashboard summarising all the various nature-related datasets held by state and semi-state bodies would be useful, especially if data was originally collected with the same infrastructure in mind, with greater coordination of metrics and metadata to facilitate more streamlined presentation and searches. Such an approach would also improve the efficiency of data collection and prevent replication.



Accessibility Challenge 2:

A lack of interoperability, with major difficulties of sharing data between different stakeholders using different systems.

Recommendation:

An amount of investment and time is needed in setting up a common system and getting buy-in from the data-gatherers and users. This will become more pressing as we collectively move toward the idea of a Public Digital Infrastructure as a “shared societal service.” See the [European Data Strategy](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en) which includes plans to organise Common European Data Spaces in areas such as Agriculture, Health, Energy and the Green Deal.* Further areas that must be addressed for these systems to be successful involve ensuring strict protocols on data privacy, security and governance and how best to engage people and organisations.

*https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en



“From a Health & Wellbeing point of view, sensitive info is necessarily subject to safeguards and confidentiality, but it’s frustrating to see data made available to big corporations but not NGOs, academics or public without jumping through hoops.”

Accessibility Challenge 3:

Obtaining data is made more difficult for some than others, for instance through charging of fees, filling out forms or long waiting periods.

Recommendations:

More open data guidelines adhered to by all stakeholders, with data delivered through simplified, aligned and centralised sources, requiring greater collaboration and coordination between large data centres and SMEs “as something can be learned from both sides” ... “maybe this is something for the tech community to lead to facilitate that collaboration.”

For an example of good practice in data accessibility see The British Ecological Society* for details of a recent initiative to provide an accessible platform for making information in grey literature (including data resources) more accessible. It also has a very powerful search engine that enables information to be tagged and searched efficiently.

*<https://www.britishecologicalsociety.org/applied-ecology-resources/about-aer/>



4. Usability

“As you collect data, think about ‘Future You’ – what might you want to use the data for in the future? And what might others want to use it for in the future?”

The above point was highlighted in one workshop presentation regarding data ‘re-use’ and how it is helpful when designing your data collection to keep in mind how you or others might one day seek to re-use the data.

Challenge 1: Lack of accurate and complete metadata at a fundamental stage.

Recommendation: Standards and procedures for the application of high-quality metadata* must be a fundamental consideration when starting out in data gathering and are essential for confident, efficient data use. Complete and accurate metadata is important for confident interrogation of spatial and non-spatial datasets. Metadata should be provided for individual Geographical Information System layers to allow for the easy interpretation of field names. A standard approach to coordinate systems was also highlighted as very important for spatial datasets.

*<https://www.dcc.ac.uk/guidance/briefing-papers/standards-watch-papers/what-are-metadata-standards>



Usability Challenge 2: Usability is often dependent on skill level and there is a lack of awareness around data governance and how community groups can use/get involved with data.

Recommendation: Thinking about your audience when collecting, designing and displaying datasets. Employing plain language in metadata is recommended where possible, and for dashboard-use instructions. It is important for open public access to demystify data from very technical terms into a common language. The National Adult Literacy Agency campaigns for public information to be put into plain language and collaborates with public bodies and organisations to provide training on this.

Another suggested way to become more user-friendly is to make your examples relatable when dealing with data useful for community groups - think about what's going to resonate with the people who you want this data to affect.



“Nothing is more frustrating than having to extract data from a pdf doc.”

Usability Challenge 3:

Data supplied in certain formats such as PDF (portable document format) is often not immediately useful e.g. if the document is secured from editing permissions. Very often no information is provided on when and how the datasets were collected, cleaned or updated.

Recommendation:

Formats such as PDFs do not fall under the definition of ‘Open Data’ - the end-user and purposes of the data (primary and secondary uses) must be considered at the earliest stages of setting out to gather data. Major public bodies should provide a spatial framework for viewing datasets. Very clear metadata procedures, following accepted standards, should be designed with non-specialist data users in mind.

Challenge 4:

Too much information can be overwhelming, for example, where there are so many layers to a dataset that it makes it difficult for the non-expert to find the information they seek.

Recommendation:

While specialised users i.e. those with experience with GIS/spatial data, may require and be very comfortable working with complex data layers, thought could be given to developing a more streamlined option for the less specialist user within the same dashboard.



Usability Challenge 5:

Mismatched data - environmental researchers and others would benefit from better coordination between some of the larger gatherers of environmental and nature-related data. For example, the INCASE project found that when using datasets from different agencies on a pilot project to map natural capital extent and condition at catchment level in Ireland, one encountered several different coordinate systems, resolutions and minimum mapping units as well as different time series for datasets.

Recommendation:

An overall data management plan should be developed for state and semi-state bodies. This plan would establish principles and protocols for gathering and sharing data and improve communication between state bodies. The plan should follow and promote the FAIR Data Principles - of findability, accessibility, interoperability and reusability. The acronym and principles were defined in the journal *Scientific Data** by a consortium of scientists and organisations.

*<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>



5. Reproducibility

Reproducibility as a theme is an amalgam of many of the other parts of the puzzle - availability, accuracy, usability, accessibility and discoverability are all essential for reproducibility. The importance of reproducibility for monitoring ongoing changes was noted in our discussions. Can you go back to collect the same data again – and if so, can you get the same results?

Ecological systems are dynamic, with numerous factors that sculpt natural history and that investigators cannot completely control. Observations may be highly dependent on spatial and temporal context, making them very difficult to reproduce, but computational reproducibility can still be achieved. Computational reproducibility often refers to the ability to produce equivalent analytical outcomes from the same dataset using the same code and software as the original study.*

*Open science, reproducibility, and transparency in ecology, [Stephen M. Powers](https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/eap.1822), [Stephanie E. Hampton](https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/eap.1822)
<https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/eap.1822>



Reproducibility Challenge 1:

- Protocols must be in place to ensure data is recorded in a standard way, however, even when these are in place, collectors don't always follow them precisely and there is a perception that this can be a particular issue of concern regarding citizen science.

Recommendation:

- Investment into the education of participants is a significant component of citizen science and ensuring reproducibility. The need for very clear instructions on how to record data is vital (e.g., for publicly sourced datasets) and it should always be pointed out to collectors that this will help improve the reproducibility of the data in future.
- Standardisation is key as it is vital that data is of good quality, especially where it will be used to inform major decisions. There are current methodologies being employed for the verification of data e.g., a team of experts is available within the National Biodiversity Data Centre to verify the quality of data being received. Data curation is a job in itself - our stakeholders felt that there was a need for a national organisation - a single reliable body - to stand over it.



A Note on Standardisation

There are likely to be some types of environmental data that are reproducible and can be standardised but many ecological surveys require bespoke approaches that may not be reproducible in another context (e.g. the placement of mammal traps depends on the surveyors' experience and assessment of what kinds of habitats and placements are likely to be effective given the vegetation and terrain).

Reproducibility is important all the way along the line from data collection to processed data - when you collect data, can the steps be reproduced? What are the cleaning steps? If data is being cleaned in different ways using the same raw data, what are the process steps taken from the data's raw format to its processed format?



Reproducibility Challenge 2:

How is data cleaned? The cleaning process entails dealing with missing data, data in the incorrect fields, data input in the incorrect boxes, *etcetera*. It is not often clear what steps are followed for the cleaning process.

Recommendation:

Protocols should be put in place at an early stage and throughout the process to record cleaning steps. There needs to be more discussion and education in general around these cleaning steps and their importance. Cleaning steps should be scripted using code which can be published as part of the metadata. Ideally a package of raw data, cleaning code and standard data product (SDP), created using the cleaning code and raw data, would be produced. In this way, the SDP could be directly reproduced from the inputs and/or the cleaning script could be modified to produce a different data product if needed.



Reproducibility Challenge 3:

Transferability of data is necessary. Changing formats must be considered - will the data stand the test of time, for example, future changes in technology? What will happen to data stored on CD ROM/hard drive/the cloud?

Recommendation:

Thought must be given to the importance of file standardisation, especially in complex databases, e.g., Excel, Access. SQL was mentioned as the standard approach for such databases now.*

Combining data can add value to the dataset, so there is a need for a seamless format in open data, to allow other agencies to use your raw data.

*<https://en.wikipedia.org/wiki/SQL>



Reproducibility Challenge 4:

Our breakout groups highlighted the need for a ‘multi-actor’ approach in designing data gathering steps to ensure engagement with all sectors involved in the process including landowners, farmers, citizen scientists, ecologists, data scientists, community groups, businesses, environmental NGOs and those working in the public sector. This can be a difficult and time-consuming process.

Recommendation:

An investment in time, staff and funding is required to ensure innovative and holistic data practices are established and well-promoted to facilitate the long-term and future health and wealth of our data ecosystem.

The final discussion was around the establishment of a national environmental data organisation which verifies all the data, with some questions about whether this could be established in Ireland to enhance reproducibility, perhaps akin to the recent development of the National Data Infrastructure Champions Group, as chaired by the CSO, which has representatives from all departments, and has been expanded to include agencies with high-value data.*

*<https://www.ops.gov.ie/actions/innovating-for-our-future/data/>



6. Discoverability

Discovering data can be difficult for two main reasons - if an unsuitable format for gathering or presenting data has been used in the first place, it may make it difficult for potential end users to discover and access it, and secondly, if no thought or effort has been put into making the wider public (or even the target audience) aware that the data exists, it may go largely undiscovered and unused.

Challenge 1: The first challenge of discoverability starts at the data gathering stage - has the correct format been established for the data?

Recommendation: At the gathering stage, we need to keep in mind the end users (and potential secondary uses) and to consider their future needs when it comes to using data and discovering it in the first place. All data collectors must consider an appropriate system of tagging for the data which will facilitate searches.



Discoverability Challenge 2:

How can we ensure people are aware that the data exists?

Recommendation:

A central signposting system within an aligned data dashboard would help to point people in the right direction. Efforts must be made through various relevant sources and media to raise awareness amongst the wider public on what data is available and how to access it. The point was raised that the time is right for leapfrogging off a general wider interest in environmental matters in order to better engage the public on nature-related data.

Challenge 3:

As it stands, even data scientists and gatherers can be uncertain as to where their data needs to go – will it end up in a silo, not seeing any major traffic? Assuming it has potential benefit across other sectors, how best to get it out there?

Recommendation:

A media campaign to promote a one-stop nature dashboard with supports to help users with visualisation, analysis of data etc. The British Ecology Society's AER* was explicitly set up to make grey literature information more easily discoverable on one platform.

*<https://www.britishecologicalsociety.org/applied-ecology-resources/about-aer/>



Conclusion

The Open Data Strategy mission statement includes “engaging with a broad community of stakeholders to promote and encourage the use of Open Data” and this was a sentiment echoed by presenters and attendees at our workshop.

This strategy envisions that Ireland become a leader in the use of Open Data and create an environment where the economic, social and democratic benefits of it are “recognised and realised”. If successfully implemented, the recommendations offered by Ireland’s nature data community will bring new opportunities for research, innovation, and greater efficiency across sectors, boosting Ireland’s green recovery by facilitating decisions that will help tackle the biodiversity crisis, while stimulating economic growth.



Appendices



**DATA4
NATURE**

Appendix 1

Breakout room prompts

Breakout room topics and guiding questions

Data Availability

- *Data availability is the process of ensuring that data is available to end users when and where they need it.*
- To what extent is Irish nature-related data readily available?
- What are the different types of end user for nature-related data?
- What are procedures are needed to ensure data is available into the future (e.g. IT, data management)?

Data Accuracy

- *Data accuracy refers to error-free records that can be used as a reliable source of information. Correct data is critical for the success of strategic decision-making.*
- How would you describe the quality of nature-related data in Ireland?
- Are there any challenges relating to the entry of data and managing errors?
- To what extent is accurate nature-related data used in decision-making in Ireland?



Data Usability

- *Data is only valuable if it's useful. Usability is the degree of ease with which data, platforms and products can be used to achieve goals. Usability assesses the level of difficulty involved in using a user interface.*
- What are the different needs of users of nature-related data?
- What processes should data publishers have in place to ensure user needs are met?
- Are there any challenges to using different data platforms e.g. Data.gov.ie, NPWS, OSI etc.?

Data Accessibility

- *Data accessibility refers to a user's ability to access or retrieve data stored within a database or other repository. Users who have data access can store, retrieve, move or manipulate stored data and fully leverage it for their individual purposes.*
- How easy is it for users to access or retrieve nature-related data?
- Can data be easily manipulated and used for research and other purposes?
- What are the challenges and opportunities relating to accessing nature-related data in Ireland?
- How can data be better synchronised across platforms e.g. scale, units, format etc.



Data Reproducibility

- *Data reproducibility refers to the degree of stability of the data when measurement is repeated under similar conditions. Ways to improve reproducibility include making raw data sets available and transparent reporting, including of negative and statistically insignificant results.*
- What is the status of reproducibility of nature-related data in Ireland?
- Are raw data readily available?
- How transparent is the reporting on data e.g. information on how the data was gathered; if, when and how it was transformed etc.

• Data Discoverability

- *Discoverability refers to the level of ease by which an individual can locate available data, detect its quality, understand its structure and lineage and gain access to it.*
- How easy is it for users to find nature-related data? Are datasets well publicised?
- Is the data properly categorised, tagged and searchable?
- What are the challenges and opportunities relating to accessing nature-related data in Ireland?



Appendix 2

Further Information on Data Dashboards

- Data dashboards are central locations that integrate data from many sources.
- Data dashboards can be hosted on data portals such as the Open Data Portal
- A data dashboard is a destination for viewing analytics. They use building blocks such as tiles to provide an intuitive 'at a glance' view for data users.
- The visual nature of a dashboard simplifies searching for data, making it an ideal platform for both analysts and non-experts such as citizen scientists to find, access and use data for different purposes.
- Examples of nature data dashboards can be found at <https://www.eea.europa.eu/themes/biodiversity/state-of-nature-in-the-eu/article-17-national-summary-dashboards>
- Data dashboards are always being modernised and improved. Designing Ireland's nature data dashboard should be done in collaboration with Ireland's data community to ensure it meets all necessary requirements.



Appendix 3

Relevant links & further reading – a list of projects and datasets referenced at our Data 4 Nature workshop

Host

www.naturalcapitalireland.com – see also www.incaseproject.com

Funders

- Department of Public Expenditure & Reform <https://www.gov.ie/en/organisation/departments/public-expenditure-and-reform/>
- Office of Public Works <https://www.opw.ie/>

Speakers

Ed Curry <http://edwardcurry.org/>

- Insight SFI Research Centre for Data Analytics <https://www.insight-centre.org/>

Yvonne Buckley Nature+, TCD <https://naturalscience.tcd.ie/natureplus/>

- Irish Research Council <https://research.ie/>
- All-Island Climate & Biodiversity Research Network <https://www.marei.ie/all-island-climate-and-biodiversity-research-network/>
- See also PlantPopNet <https://www.plantpopnet.com/>



Speaker links ctd: Gemma Weir

- National Parks & Wildlife Service www.npws.ie/maps-and-data
- See also <https://www.npws.ie/maps-and-data/open-data-policy>
- NPWS Site Specific Conservation Objectives:
<https://dahg.maps.arcgis.com/apps/webappviewer/index.html?id=63b6a14f5b164b289ad87048f71532b8>
- National Ecosystem & Ecosystem Services Mapping Pilot:
<https://dahg.maps.arcgis.com/apps/MapSeries/index.html?appid=cb5040a4a19645b6b424bed940c54fff>
- NPWS Data Request Form <https://www.npws.ie/maps-and-data/sensitive-data-access>
- NPWS publish at: <https://data.gov.ie/>
- Irish Spatial Data Exchange <http://www.isde.ie/>
- Inspire Geoportal <https://inspire-geoportal.ec.europa.eu>
- NPWS data-sharing initiatives:
- Geohive <http://map.geohive.ie>
- Natura 2000 Network Viewer <https://natura2000.eea.europa.eu/>
- Central Statistics Office www.cso.ie



Speaker links ctd. Liam Lysaght

- National Biodiversity Data Centre www.biodiversityireland.ie
- Compass Informatics <https://compass.ie/>

Tony Brew, Office of Public Works www.opw.ie

- <https://www.gov.ie/en/publication/b15dd0-technical-specifications-and-guidance-notes/>

Paddy Morris, EPA www.catchments.ie

Lisa Coleman, INCASE www.incaseproject.com

Luke Binns www.SmartDublin.ie

- Dublinked - <https://data.smartdublin.ie>
- European Open Data Portal <https://data.europa.eu/>
- Operandum <https://www.operandum-project.eu/>
- Dublin Dashboard www.dublindashboard.ie
- Access Earth <https://www.access.earth/>
- Hale & Hearty www.haleandhearty.ie



Speaker links ctd. Ainhoa Gonzalez

- ESM Webtool <https://enviromap.ie/>
- All-Ireland Research Observatory <https://airo.maynoothuniversity.ie/>
- Geohive <https://geohive.ie/>
- Ordnance Survey Ireland <https://www.osi.ie/>
- Office of the Planning Regulator <https://www.opr.ie/>

Rowan Fealy & Tim McCarthy, Terrain-AI <https://terrainai.com/>

- Derilinx Open Data Webinar <https://derilinx.com/webinar-open-data-impact-series-vi-climate-change/>

Kevin Doolin, Demeter Project www.h2020-demeter.eu

- Walton Institute https://www.wit.ie/research/centres_and_groups/tssg

Kieran Sullivan, Agri-Discrete <https://waltoninstitute.ie/projects/agridiscrete>

Q&A

- Open Litter Map <https://openlittermap.com/>

