NERC COVID-19 Hackathons: Data Resources

The NERC COVID-19 Hackathons set out a number of key societal challenges related to the COVID-19 Pandemic, to be addressed over the duration of the Hackathon events. In these challenges, participants will draw together and manipulate a range of datasets, for example from the physical and environmental sciences, health and epidemiology, social science. We leave the choice for the specific datasets to you to select to help deliver your specific Digital Solution. There is a vast array of datasets now made available, and your solution may use any data sources as you see fit. The purpose of this document is to signpost you to some of the more significant data resources. Most importantly, NERC, as part of UKRI, are making the collections of datasets held by the NERC Environmental Data Service available to participants. Advice on that is provided below:

Accessing the NERC Environmental Data Service (EDS)

NERC operate a network of environmental data centres (https://nerc.ukri.org/research/sites/data/) that provide a focal point for NERC's scientific data and information. These centres hold data from environmental scientists working in the UK and around the world. The data centres are responsible for maintaining environmental data and making them available to all users, not just NERC researchers but others from science, commerce, government and education as well as the general public. NERC supports five data centres covering a range of **discipline** areas:

- British Oceanographic Data Centre (BODC) (**Marine**)
- Centre for Environmental Data Analysis (CEDA) *which includes:*
    - British Atmospheric Data Centre (BADC) (**Atmospheric**)
    - NERC Earth Observation Data Centre (EODC) (**Earth observation**)
    - UK Solar System Data Centre (SSDC) (**Solar and space physics**)
- Environmental Information Data Centre (EIDC) (**Terrestrial and freshwater**)
- National Geoscience Data Centre (NGDC) (**Geoscience**)
- Polar Data Centre (PDC) (**Polar and cryosphere**)

In addition, NERC contracts the Archaeology Data Service to manage and make available data arising from NERC funded research in Science Based Archaeology.

The NERC Data Catalogue Service provides an integrated, searchable catalogue of the data holdings of NERC's data centres. It can be used to find information on what data the NERC data centres hold and how to access these data.

For accessing NERC datasets, the following Q&A should help:

How can I access the data?

The NERC Data Catalogue Service (https://data-search.nerc.ac.uk) provides a central point of access to search NERC's data holdings and information products. An alternative approach is to visit https://nerc.ukri.org/research/sites/data and follow the links to the individual domain NERC Data Centres noted above.

Who can access the data?

For the specific purpose of the hackathons, anyone can apply for access to information and data held by NERC. The relevant Data Centres will be able to provide more information.

What can I do with it?

For the specific purpose of the hackathons, the NERC data can be used for any purpose and is supplied under the Open Government Licence.

Note that there is held within the EDS some data from third parties. This is released only under licence terms. A discussion with the relevant Data Centre will help provide appropriate guidance.

Is there any charge?

For the specific purpose of the hackathons, access to the NERC data shall be provided free of charge. However, that there may be a charge for some information products containing third party data – something to discuss with the Centre staff.

What if I can't find what I'm looking for?

Contact the relevant Data Centre for specific data enquiries. They have specialist domain knowledge on suitability of the NERC data sets.

Other Key Environmental Data Sources

European Space Agency (ESA) and the Copernicus Data Service
Another key source of environmental data is the Copernicus Data Service. Of particular use for example could be the ERA5/ERA5t data that are provided by

ECMWF. Those are available through the [Copernicus Data Service](). The NERC CEDA service have a limited selection of those data so it may be advised to make use of the data from the CDS as the primary source rather than the CEDA archive at present. This data is made available under the [following licence](). This means that Hackathon participants can make good use of them. Note there is an API to access the data from the CDS too.

For a summary of the ESA satellite platforms (Copernicus) and some ideas, see – https://www.esa.int/Applications/Observing_the_Earth/COVID-19_how_can_satellites_help

Met Office

The UK Met Office offer the COVID-19 data lab – https://covid19-response.informaticslab.co.uk/hub/login

To gain access to the data platform is free but you need a GitHub account and then that is linked to the logon for the site. There are instructions there to help you. At the moment there is data from 1st Jan this year as focus was on data from during the pandemic. It includes:

- Global gridded hourly surface meteorological data – this is on a 10km grid from our global predictions
    - UK gridded hourly surface meteorological data – this is on a 1.5 km grid from our UK predictions
    - This data is also aggregated up to daily gridded average, min, max values. Done to reduce data volume and as lots of other people's data is on coarse temporal scale and so forcing them to use hourly is not very 'user friendly'
    - The Met Office have also further aggregated the data spatially onto health reporting regions, i.e. 'get average temperature for a day for a county'. Again, this is because a lot of the health data is region based. This is done based on areas defined in shape files and easy to do for any areas. Example python code to do this is also included on the platform.
    - The Met Office have also added air quality gridded model data now and are keen to explore other data.
    - The platform uses [Pangeo]() and the Met Office have set it up on Microsoft Azure with some compute power behind it and so this allows people to run analysis using a range of tools and languages (Python and R are fully set up).

Users have not only been from the health area but also form behaviourally science and policy areas, i.e. wanting to look at adherence to lock down rules in relation other weather.

COVID-19 Data Sources and Resources:

In addition to the NERC environmental datasets outlined above, there is also a broad range of publicly accessible COVID-19 related data resources which may aid your work.

Listed here are some of these principal sources:

- UKRI Coronavirus Hub - https://www.ukri.org/research/coronavirus/
- Data.gov.uk COVID-19 dashboard - https://coronavirus.data.gov.uk
- NHSX - https://www.nhsx.nhs.uk/
- NHS Pathways Coronavirus (Covid-19) dashboard - https://digital.nhs.uk/data-and-information/publications/statistical/mi-potential-covid-19-symptoms-reported-through-nhs-pathways-and-111-online
- COVID-19: track coronavirus cases - https://www.gov.uk/government/publications/covid-19-track-coronavirus-cases
- ONS Latest data and analysis on coronavirus (COVID-19) in the UK and its effect on the economy and society - https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases
- CORD-19 (White House, USA) - https://pages.semanticscholar.org/coronavirus-research
- CoViz - https://coviz.apps.allenai.org/
- CSSEGISandData/COVID-19 - https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series
- Novel Coronavirus (COVID-19) Cases Data - https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases
- Google
    - Mobility reports - https://www.google.com/covid19/mobility/ (explanation link)
    - Covid-19 information and resources - https://www.google.com/covid19/
- European Centre for Disease Prevention and Control - https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide
- ESA satellite platforms (Copernicus) - https://www.esa.int/Applications/Observing_the_Earth/COVID-19_how_can_satellites_help

- Open access epidemiological data from the COVID-19 outbreak - Open COVID-19 Data Curation Group - https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30119-5/fulltext#
- Nature Research Data Support service for COVID-19 - https://www.springernature.com/gp/researchers/campaigns/coronavirus/research-data
- CityMapper, City mobility - https://citymapper.com/cmi
- Namara - https://how-to.namara.io/#introduction
- GitHub (various sources) - https://github.com/topics/covid-19

Specific Useful Key Document Sources

You may find it helpful to review the following further documents as you develop your plans:

- WHO. COVID 19: Public Health Emergency of International Concern (PHEIC). Global research and innovation forum: towards a research roadmap [Link]
- SCEID - National Academies of Science, Engineering, Medicine. Standing Committee on Emerging Infectious Diseases and 21st Century Health Threats: Virtual Meeting 1 [Link]
- Artificial Intelligence against COVID-19: An Early Review [Link]
- Conticini, E., Frediani, B., Caro, D. (2020) Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? Environmental Pollution. https://www.sciencedirect.com/science/article/pii/S0269749120320601
- Environmental Justice Commission (2020) Faster, further, fairer: Putting people at the heart of tackling the climate and nature emergency. Online at https://www.ippr.org/research/publications/faster-further-fairer
- Greenpeace (2020) A Green Recovery: How We Get There – Greenpeace sets out plans for a greener, cleaner, fairer future. Online at https://www.greenpeace.org.uk/news/a-green-recovery-how-we-get-there-greenpeace-sets-out-plans-for-a-greener-cleaner-fairer-future/

Tools source

The following tools may be of use to you in developing the digital solutions you put forward:

- ArcGIS Online - https://www.esri.com/arcgis-blog/products/bus-analyst/real-time/latest-county-level-covid-19-data-in-arcgis-business-analyst/
- https://www.visualcapitalist.com/7-best-covid-19-resources/
- https://www.thecoronavirushackathon.com
- Paraview, large scale data visualisation - https://www.paraview.org/overview/
- D3, data driven documents - https://d3js.org
- Matplotlib, visualisation with Python - https://matplotlib.org
- Seaborn, statistical visualisation - https://seaborn.pydata.org
- Plotly, front-end for AI and ML - https://plotly.com

- Google Charts, interactive charts & data tools - https://developers.google.com/chart
- ChartJS, HTML5 charts - https://www.chartjs.org

<end>