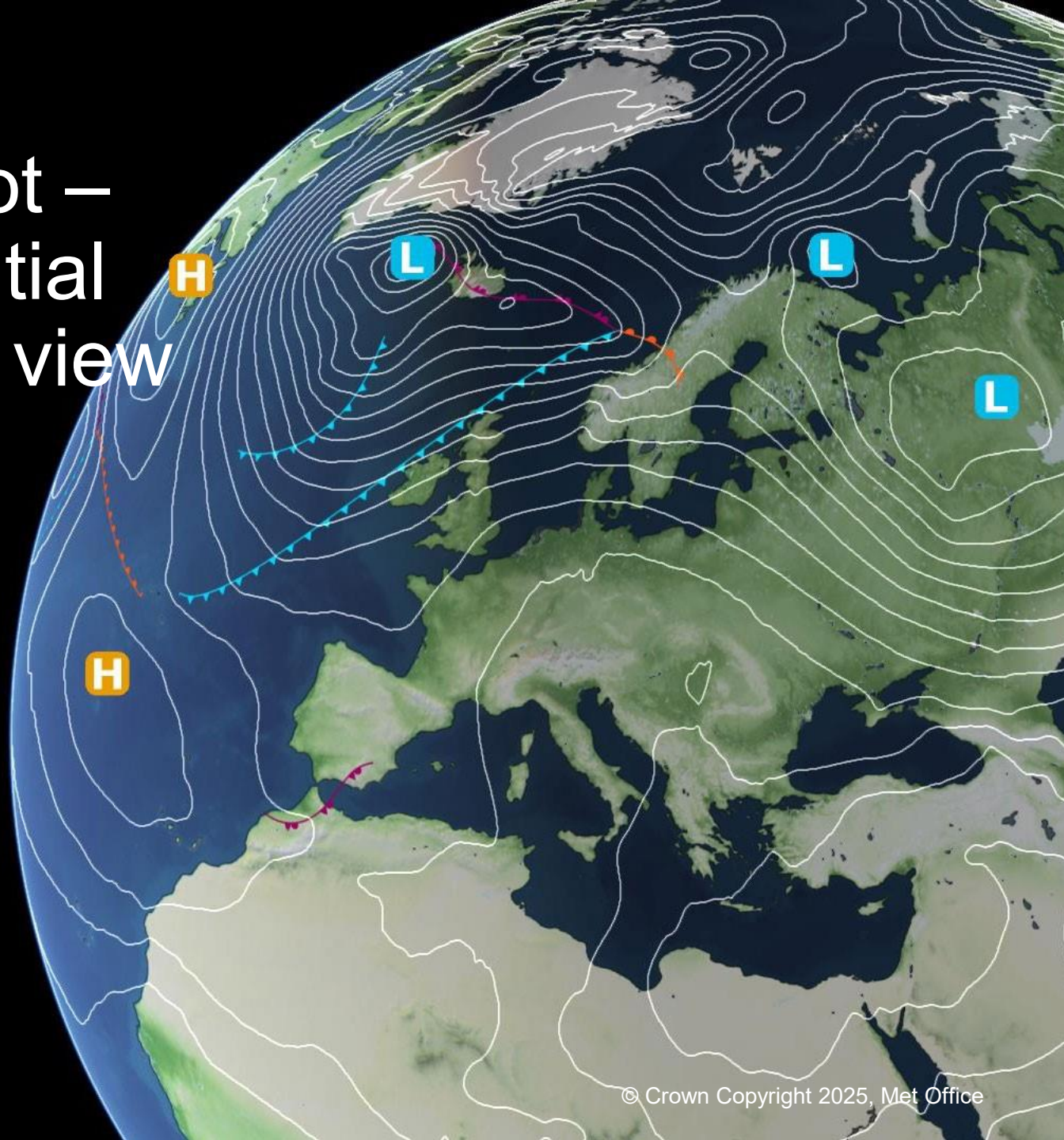
Avoiding the Gordian Knot – the challenges and potential from an Archives point of view

UK Met Office National Meteorological Archive data rescue activities

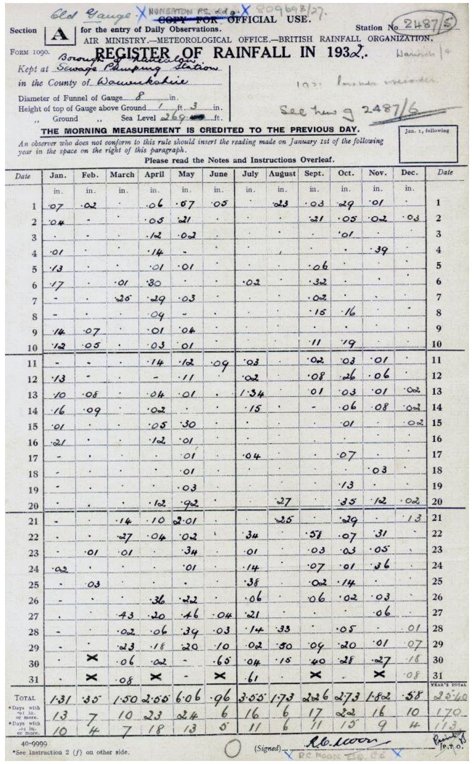


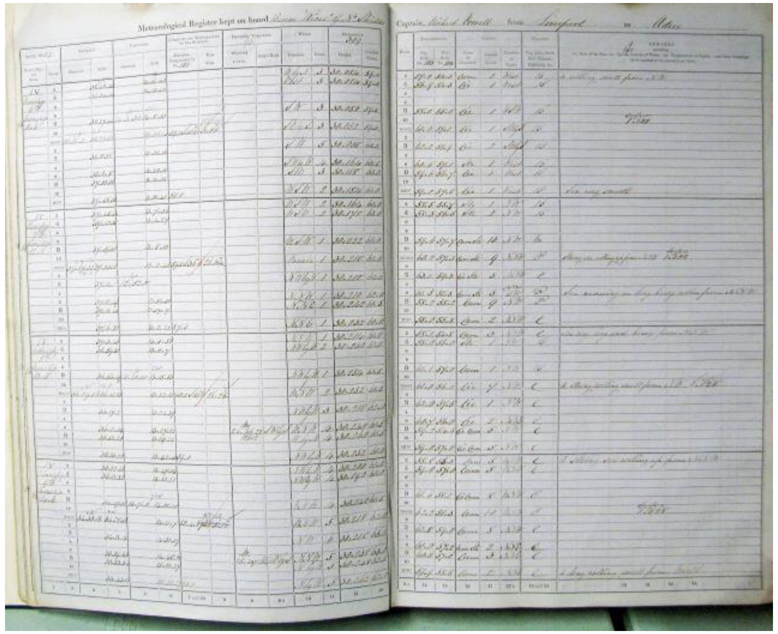
NMLA holdings

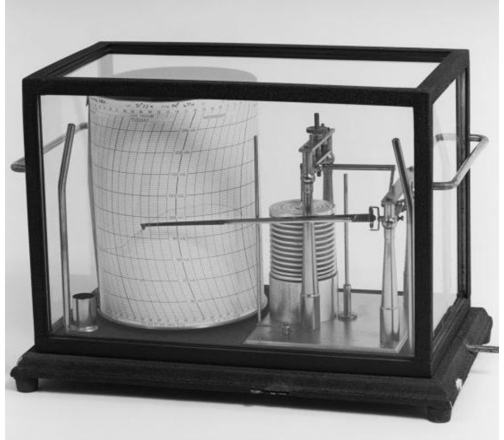
- National Memory of the Weather
- Data
- Metadata
- International data which places the UK data in context
- Earlier data including private weather diaries

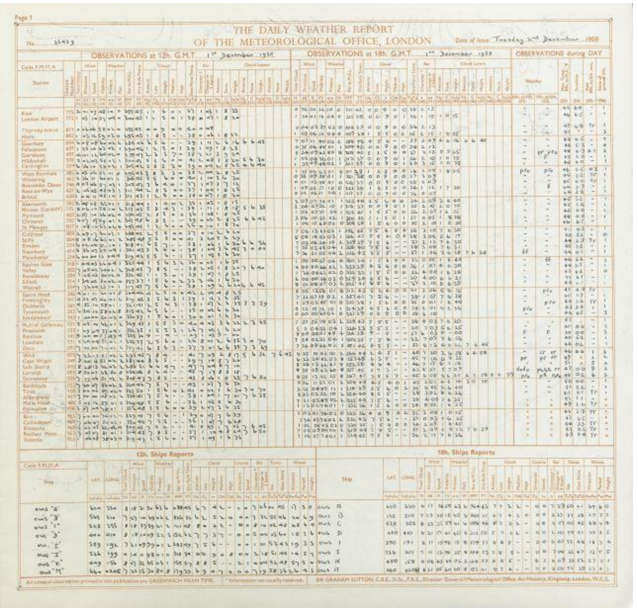
- In recent years there has been significant interest in recovering information from a number of our most important and long running collections.

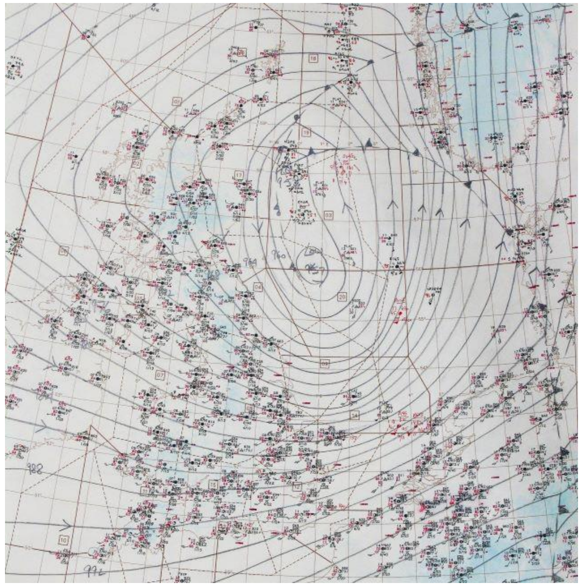
- www.metoffice.gov.uk/research/library-and-archive

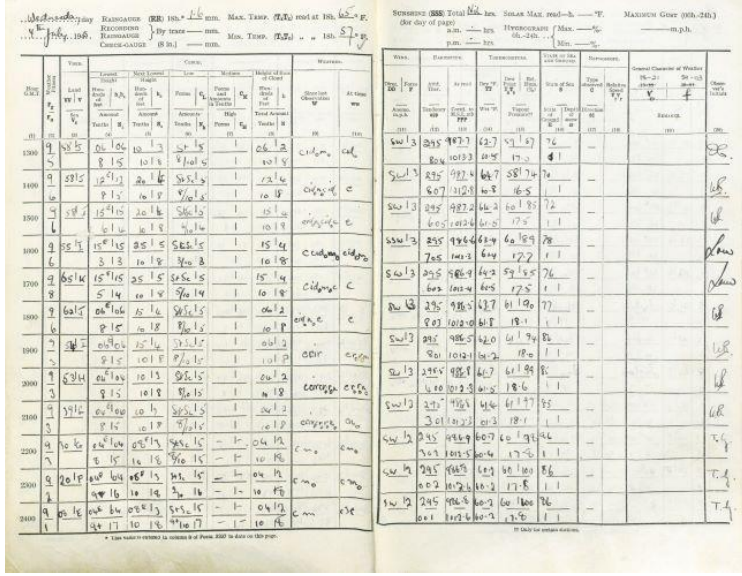

FOR OFFICIAL USE.
AIR MINISTRY.—METEOROLOGICAL OFFICE.—BRITISH RAINFALL ORGANIZATION.
REGISTER OF RAINFALL IN 1932.
THE MORNING MEASUREMENT IS CREDITED TO THE PREVIOUS DAY.
Please read the Notes and Instructions Overleaf.

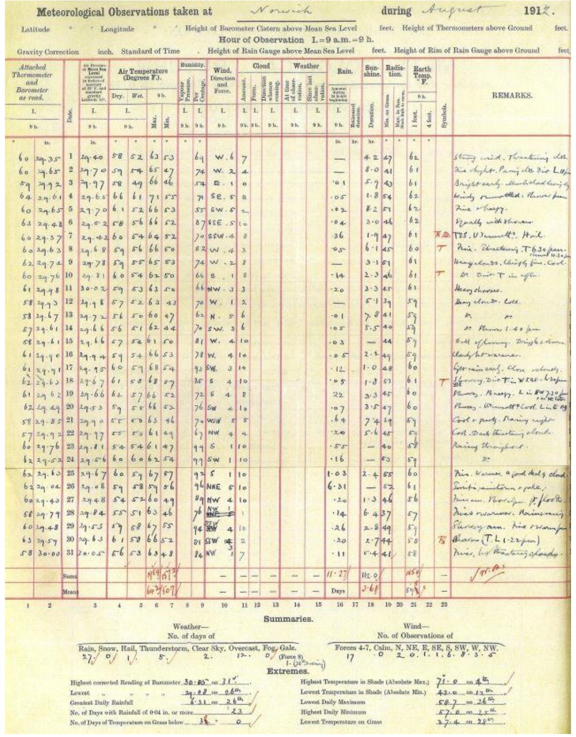





THE DAILY WEATHER REPORT
OF THE METEOROLOGICAL OFFICE, LONDON





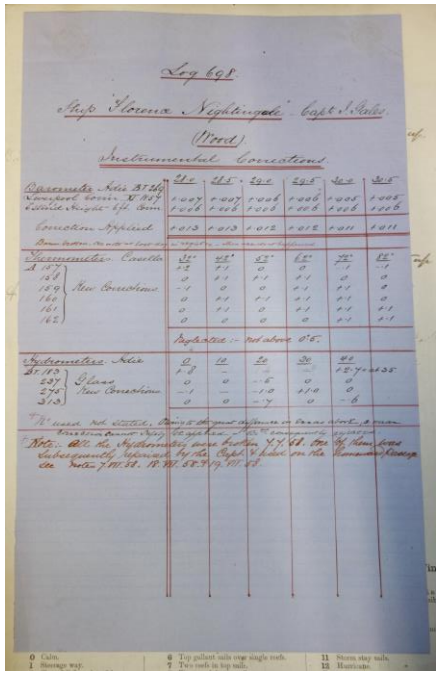

Meteorological Observations taken at Norwich during August 1912.
Summaries.
Extremes.

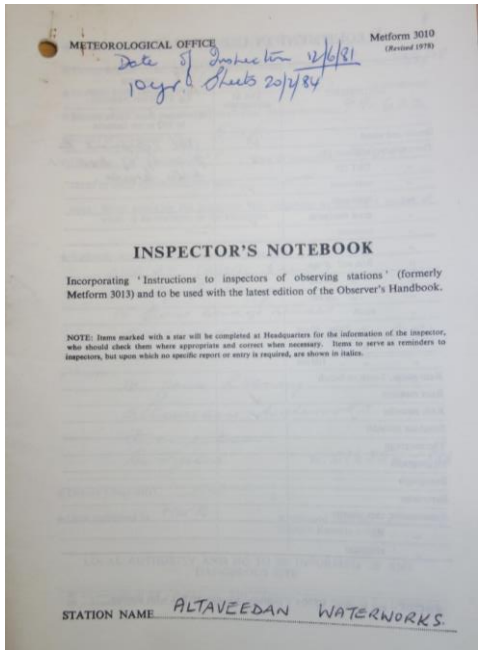
Met Office

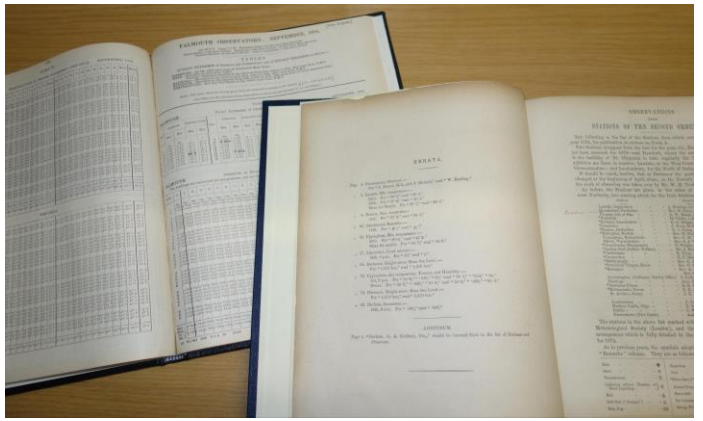
Recovering metadata

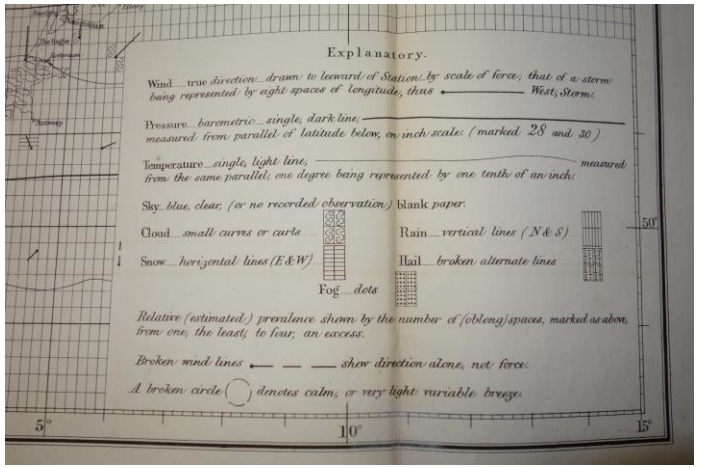
Metadata can be found in many places in an archive

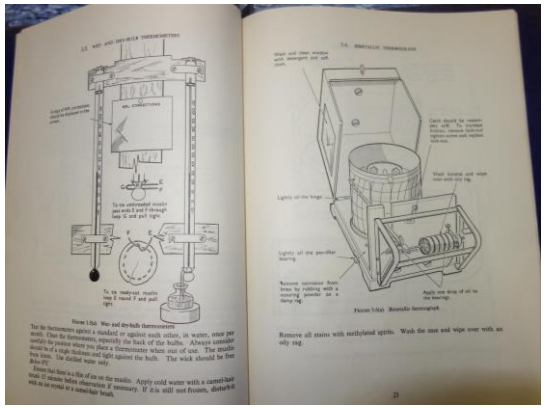
- Station Inspection documents
- On the data itself
- Covers and inside pages
- Manuals, keys decodes and publications

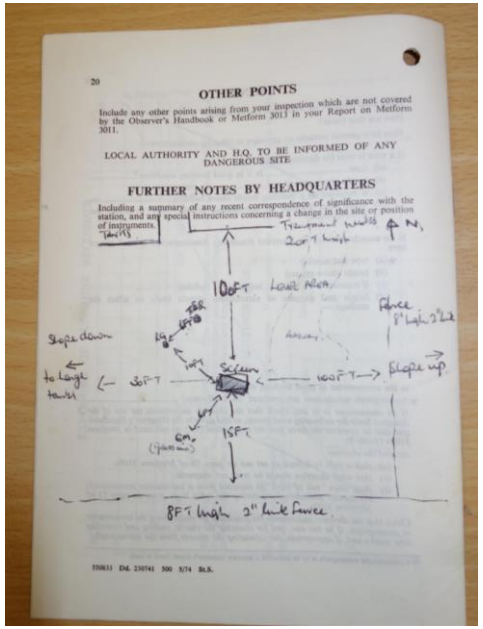


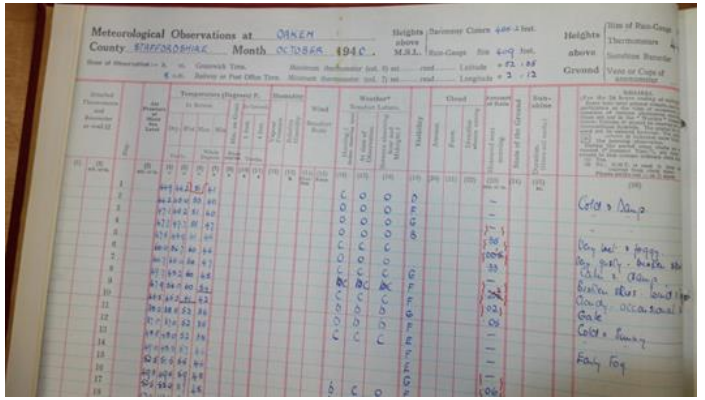












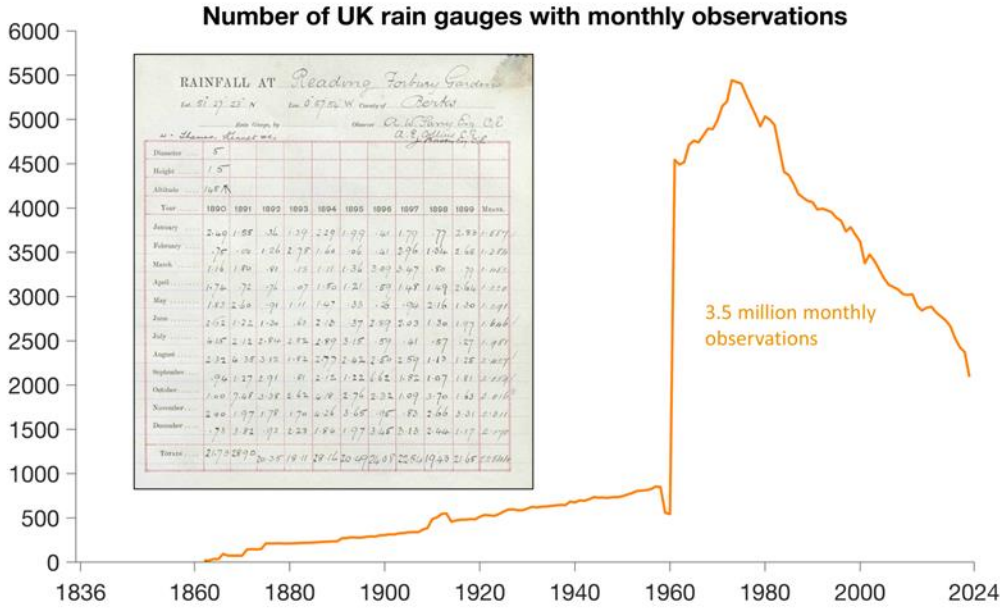
Met Office

Engagement with the data rescue community (so far!)

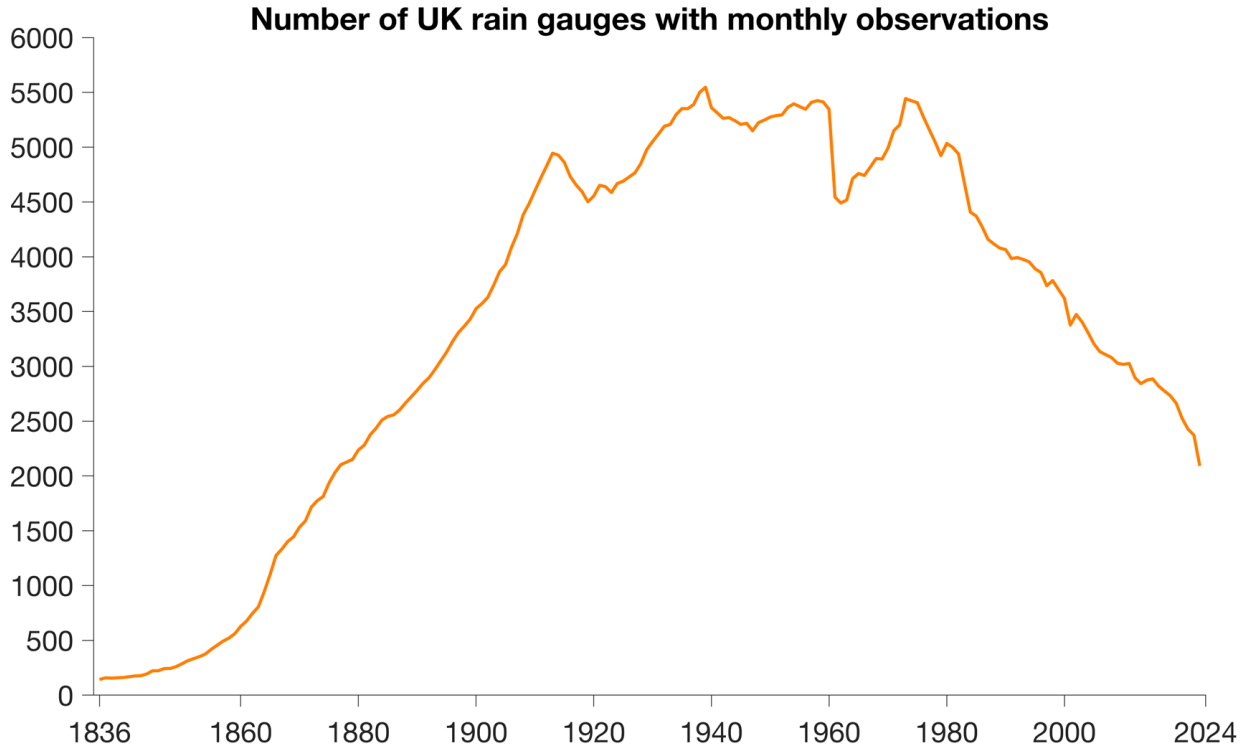
- We work with the UK Met Office, the University of Reading, the British Antarctic Survey and global communities such ACRE and GloSAT
- There are literally tens of millions of observations in the paper archive which are not in the electronic data sets. These generally stretch from the 1950s backwards, often well into the 1800s.
- Around a decade ago we started to scan our data to make it more accessible for researchers and it is already having impacts.

Using data rescue to inform climate science UK examples

- Rainfall rescue and the blended rainfall database – monthly rainfall
- Storm Reanalysis – 2nd order stations, hourly observations, daily weather reports
- Creating new surface temperature datasets – international climate returns, international published data, ship logs
- Creating new datasets, extending and gap filling – climate returns, agricultural returns, 2nd Order Stations

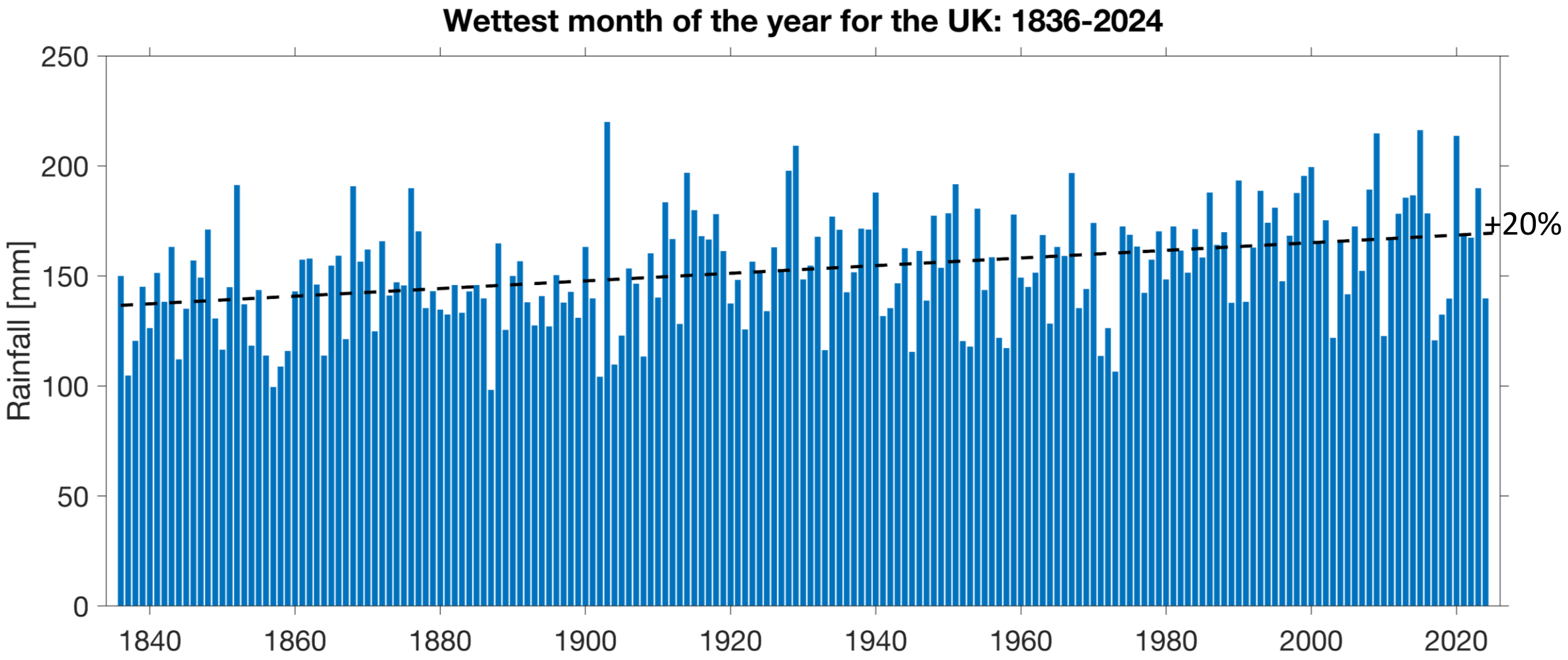


- 16,000 volunteers made invaluable contributions to science
- Detailed UK rainfall records extended backwards to 1836
- More data available for 1880s than now!



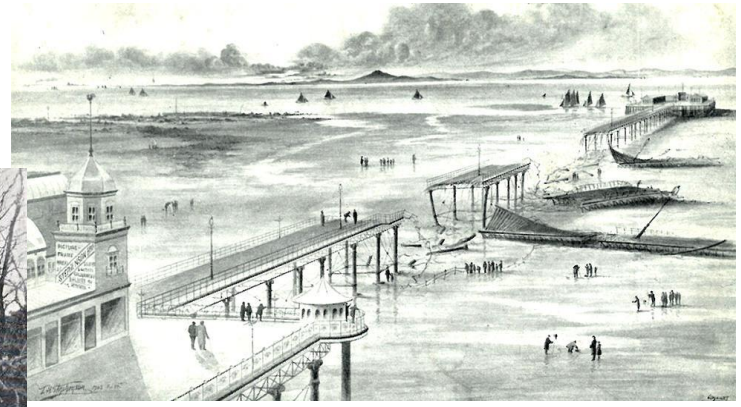
Autumn and Winter months are getting wetter….

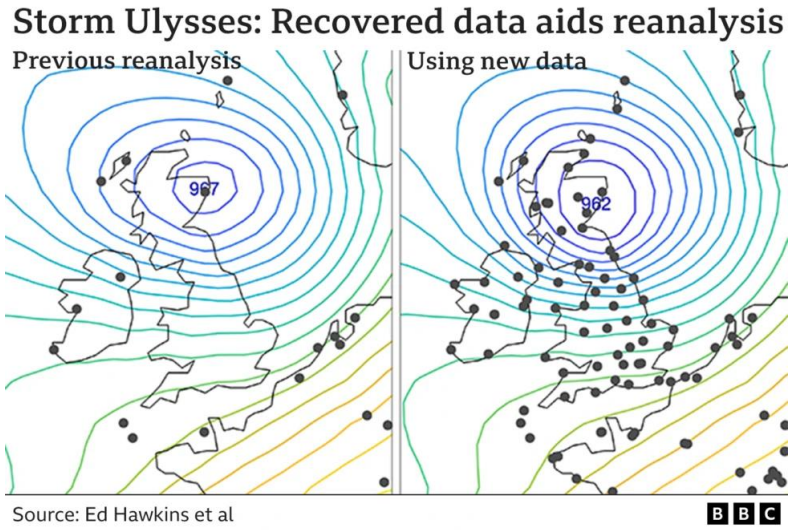
(but not much change in summer)

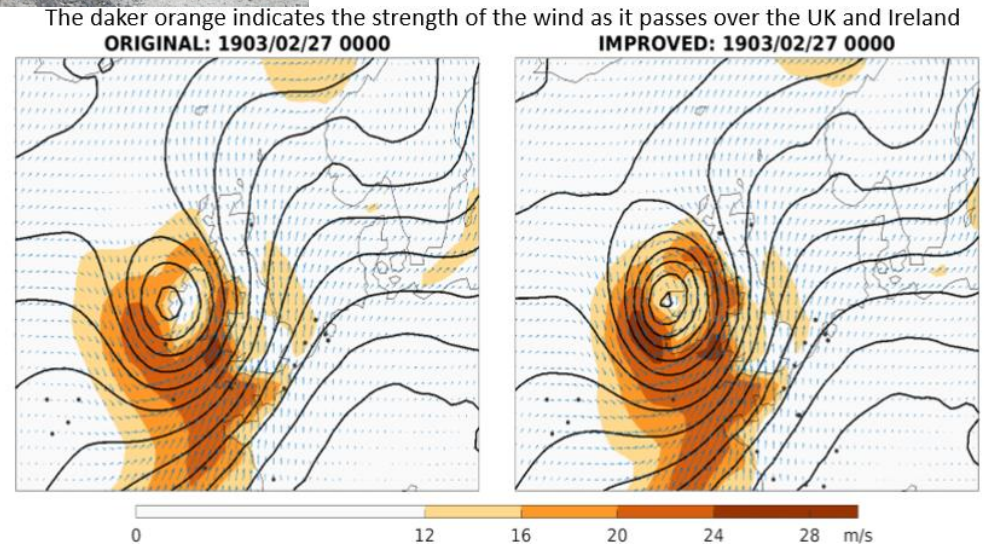


Storm reanalysis – Storm Ulysses 1903

- Reanalysis gives an accurate picture of this storm and its impacts and enables us to test and trust our models
- We can then change the starting parameters to match current climate and see how the storm would impact us now, or under future predicted scenarios
- Helps to prepare for impacts of extreme weather







Challenges for archives right now

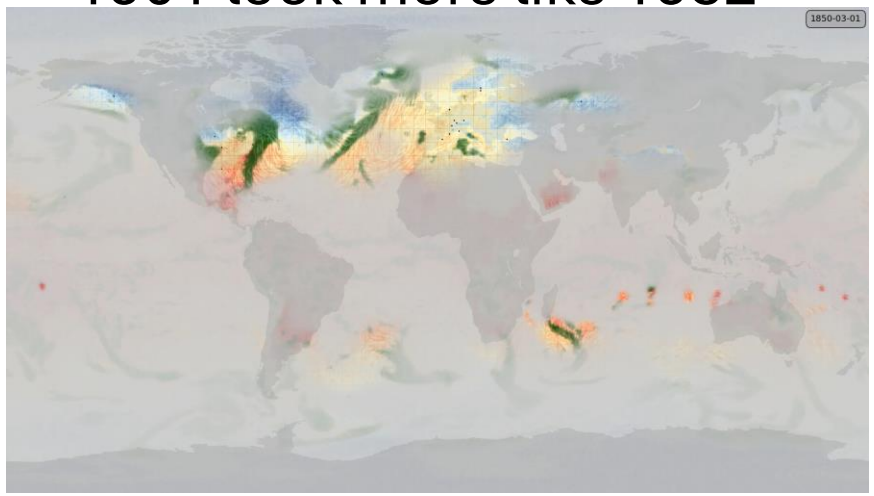
- Understanding the best collections to tackle and considering the breadth of our audiences
- Funding – scanning is not cheap and keying is generally beyond available resource. Archival storage also comes at a cost and the more we use, the more it costs!
- Managing the resulting scans – these take considerable time to download, QC, and store. We priorities making the scans easily available as this is critical for climate scientists.
- Long term preservation – ensuring the files are in long standing formats and preserved in an archive storage system which can render and migrate them to new formats when necessary.

Challenges for archives in the next 5 years

- AI – this is the only reasonable solution to the need to key millions of observations and make them fully available for climate science
- Right now there isn't a workable AI tool but things are changing at speed – within a few years there will be a trustworthy and affordable solution
- We need to be ready for that – with strategy for how to spread available funds between scanning and keying, with plans for sensible format types and file naming and with plans for how the keyed observations can be safely and permanently stored….new databases.
- We need to think about both the cost and the carbon foot print of this and ensure we are rescuing and keying useful and useable data ie. datasets which can be included in existing databases and with sufficient metadata to be trusted and used in modelling etc.
- Without thinking about this we just create a gourdian knot of numbers without context which may not even survive the next 10 to 20years if we haven't thought about formatting and storage.

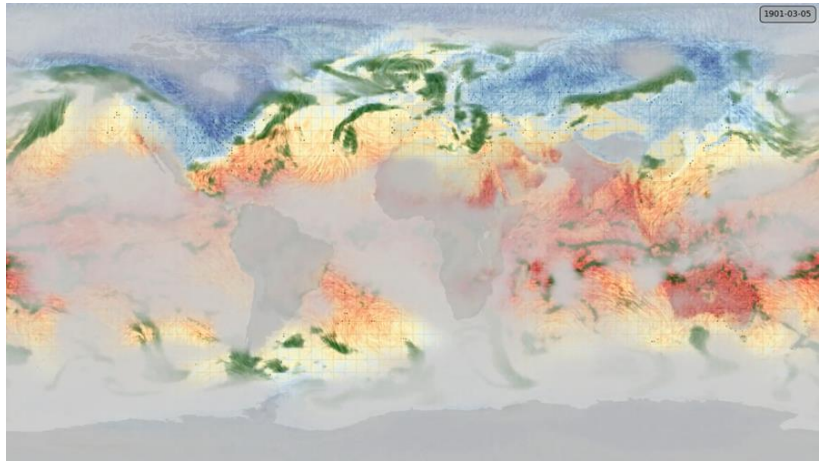
Is it worth it?!

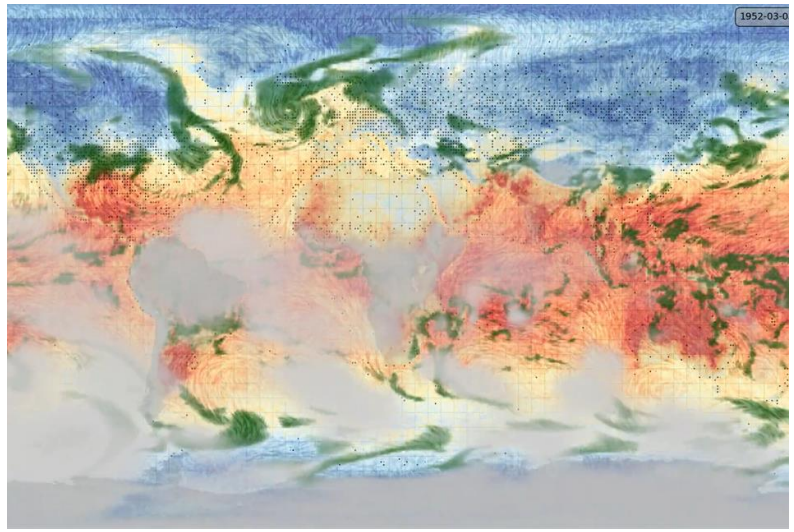
- Remaining relevant - managing the information properly and keeping it preserved means Archives can continue to be actively engaged in supporting climate science and remain trusted sources of data for partners around the world
- And every observation we can recover that is eventually fed into a model – be it 20CR, ERA5 or something else helps to clear the fog of uncertainty and make 1850, or at least 1901 look more like 1952



1850

1901





1952

Weather and fog 1850-2015 — Tests and analyses of version 3 of the Twentieth Century Reanalysis with thanks to Philip Brohan