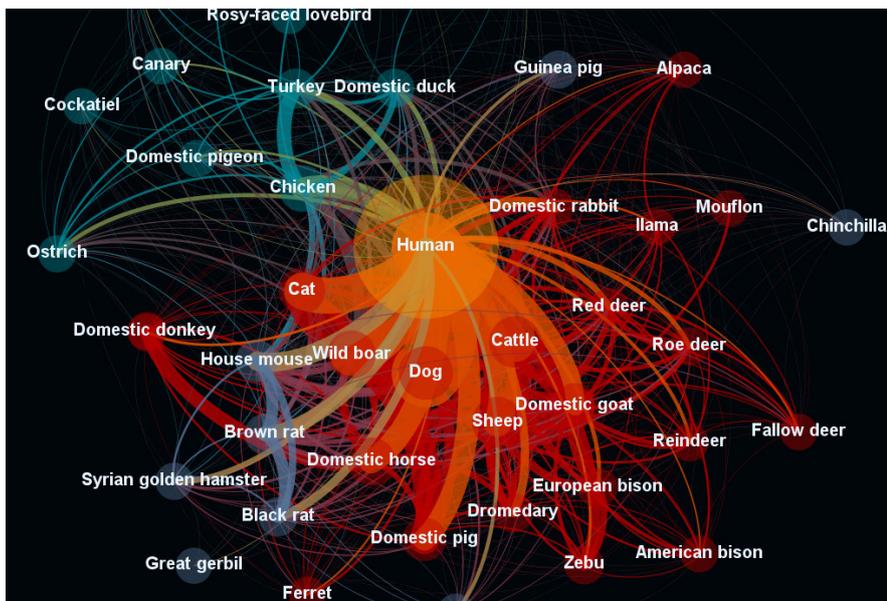


# Using big data to map and defeat disease



Mapping the relationship between domestic animal and human host species. The size of nodes represents the number of pathogen species found for each host; the arrows linking nodes show the number of pathogen species shared by each. Copyright: Dr Maya Wardeh

## Dr Marie McIntyre:

“EID2 is matchless in scale and has the capacity to hold data on all known human, animal and plant pathogens.

“There are already tens of millions of sequence uploads to look at, and millions more are added every year. EID2 has the capacity to become a comprehensive, definitive source of pathogen and disease information.”

## Scientists funded by BBSRC are building the world’s most comprehensive animal and human pathogens database to prevent and tackle disease around the globe.

The Enhanced Infectious Diseases (EID2) database was created by the Liverpool University Climate and Infectious Diseases of Animals (LUCINDA) team.

Mapping the relationships between human and animal diseases, their hosts, pathogens and transmission routes helps us learn what disease risks exist in a population or region, and how best to tackle them.

It has been used to trace the history of diseases, to predict the effects of climate change on pathogens, to produce maps of which diseases are most likely in an area and to categorise complex relationships between the carriers and hosts of numerous pathogens.

Epidemiologist Dr Marie McIntyre, one of the EID2 team, said: “The database is matchless in scale, and has the capacity to hold data on all known human, animal and plant pathogens, when detailed information becomes available.

“After scientists have sequenced part or all of a pathogen’s DNA or RNA, they usually upload the sequence to public databases, and include information on where, when and from which host the pathogen was obtained.

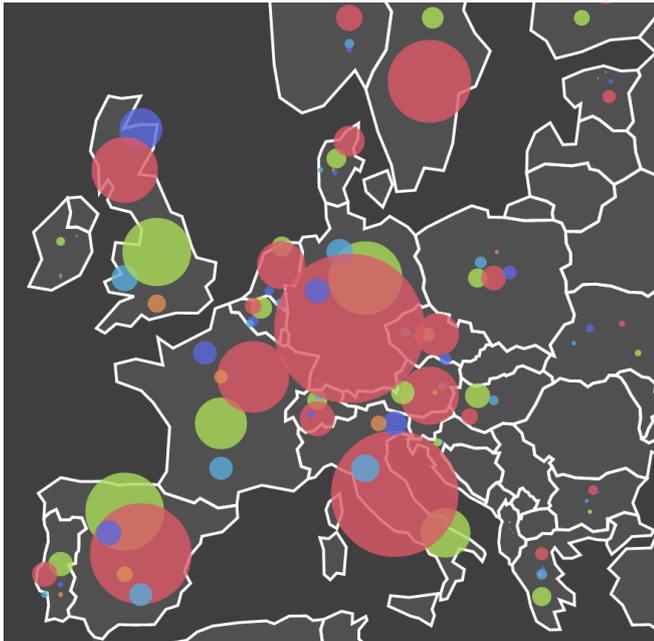
“There are already tens of millions of sequence uploads to look at, and millions more are added every year. EID2 has the capacity to become a comprehensive, definitive source of pathogen and disease information.”

## Key facts

**99 %** EID2 information extraction algorithms are more than 99% accurate.

The database is regularly updated to be the most comprehensive source of linked host-pathogen-location information.

**42M** EID2 is built from more than 42 million DNA/RNA sequences and describes 2,595 species of pathogens.



The number and types of pathogens found in EU countries. The size of circle is proportional to the number of pathogen species, and the colours represent types of pathogen. Copyright: Dr Maya Wardeh

The database is open-access to help scientists researching all sorts of subjects.

Dr McIntyre said: "EID2 is useful because it gives access to sets of information on pathogens which have, until now, been difficult to acquire. It gives instant access to the raw data from which this information is built."

Disease mapping is an important area where EID2 is a valuable tool because only around four percent of clinically-important diseases in humans have been mapped.

EID2 can quickly build maps for human, animal and crop diseases, producing country-by-country, region-by-region or even county-by-county profiles of factors affecting disease; meaning regions can prepare to avert or manage outbreaks.

Dr McIntyre added: "EID2 can also quantify the interactions between pathogens and their hosts. We know that humans originally acquired about 60% of our pathogens from animals, but we don't know which animals. Once we have a clear picture of the pathogen species found in domestic and wild animal hosts, we will be able to study the possible routes by which pathogens make it into human populations.

"Is a new pathogen of, say, mice going to reach humans because they interact with us in our houses, or will it be

## Contact

Name: Dr Marie McIntyre  
 University of Liverpool  
 Tel: 0151 795 8307  
 Email: k.m.mcintyre@liv.ac.uk

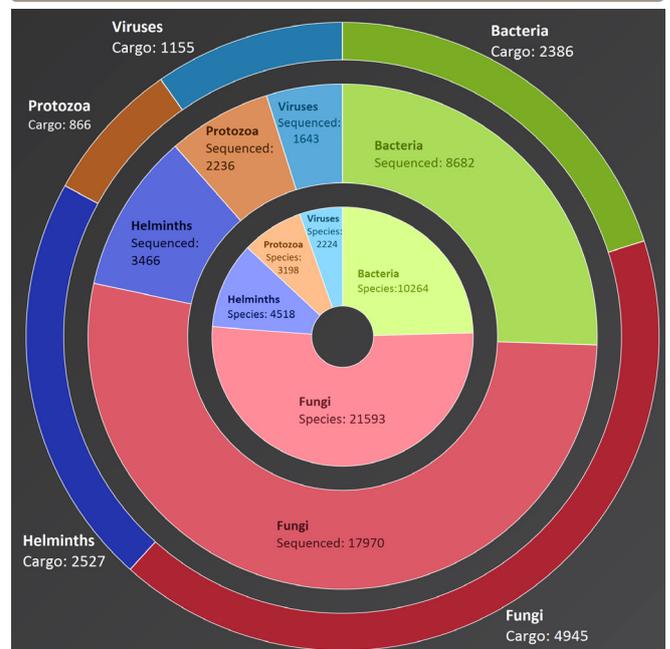
via the cats that eat them? This kind of information is ordinarily very difficult to acquire because it requires knowledge of which hosts are affected by diseases and vice-versa, but also how often these infection events occur, or if they have ever happened."

The EID2 team plans to use data for risk analysis, predicting where and in which species certain diseases are most likely to occur, and producing estimates of where diseases can occur based on environmental data such as climate, demographics and vegetation.

## About BBSRC

The Biotechnology and Biological Sciences Research Council (BBSRC) invests in world-class bioscience research and training on behalf of the UK public. Our aim is to further scientific knowledge, to promote economic growth, wealth and job creation and to improve quality of life in the UK and beyond.

Funded by Government, BBSRC invested over £484M in world-class bioscience in 2013-14. We support research and training in universities and strategically funded institutes. BBSRC research and the people we fund are helping society to meet major challenges, including food security, green energy and healthier, longer lives. Our investments underpin important UK economic sectors, such as farming, food, industrial biotechnology and pharmaceuticals.



This depicts pathogen species that EID2 has information on. The innermost circle represents the number of species listed within the NCBI taxonomy database in the major groupings containing pathogens. The middle circle is those for which sequences and metadata potentially describing their host origin are available in the NCBI Nucleotide database. The outer circle represents pathogen species for which data has been captured about the hosts in which they occur within the EID2 database itself (denoted as 'Cargo'). Copyright: Dr Maya Wardeh