Challenges and opportunities of sharing animal health data for research and disease management: a case study of bovine tuberculosis

A. Mitchell* (1), N. Alexander (2), J. Ellerbeck (1), G. Enticott (3), P. Hogarth (1), A. Prosser (1), L. Lambert (1), D. Hackett (4), N. Tait (2), J. Tiller (1), P. Upton (1) & W. Wint (2)

(1) Animal and Plant Health Agency, Weybourne Building, Woodham Lane, Addlestone, KT15 3NB, United Kingdom

(2) Department of Zoology, Environmental Research Group Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom

(3) School of Geography and Planning, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff, CF10 3WA, United Kingdom

(4) Department for Environment, Food and Rural Affairs, Nobel House, 17 Smith Square, London, SW1P 3JR, United Kingdom

*Corresponding author: andrew.mitchell@apha.gov.uk

Summary

The sharing of animal disease data should be encouraged. The analysis of such data will broaden our knowledge of animal diseases and potentially provide insights into their management. However, the need to conform to data protection rules in the sharing of such data for analysis purposes often poses practical difficulties.

This paper sets out the challenges and the methods used for the data sharing of Animal Health data in England, Scotland, and Wales – Great Britain (GB) using bovine tuberculosis (bTB) data as a case study. The data sharing described is undertaken by the Animal and Plant Health Agency on behalf of the Department for Environment, Food and Rural Affairs and the Welsh and Scottish Governments. It should be noted that Animal Health Data are held at the level of GB (rather than the

United Kingdom – which includes Northern Ireland) as Northern Ireland's Department of Agriculture, Environment and Rural Affairs has its own separate data systems.

Bovine tuberculosis is the most significant and costly animal health problem facing cattle farmers in England and Wales. It can be devastating for farmers and farming communities and the control costs for taxpayers in GB are over £150m a year.

We describe two methods of data sharing – first, where data are requested and delivered to an academic institution for epidemiological or scientific analysis, and second, where data are proactively published in an accessible and meaningful way. As an example of the second method, we describe the free-to-access website (https://ibtb.co.uk) where bTB data are published for the benefit of the farming community and veterinary health professionals.

Keywords

Animal health data – Bovine tuberculosis – Data sharing – Disease management – ibTB – Information bovine tuberculosis – UK GDPR – United Kingdom General Data Protection Regulation.

Background

The advantages to be gained from the sharing of data have long been recognised. In 2013, the Department for Environment, Food and Rural Affairs (Defra) – as part of a wider United Kingdom (UK) Government initiative – published its 'Open Data Strategy' [1], with a key principle that data should be open by default. The hope across Government was that such a strategy, delivering access to Government-held data for both individuals and companies, would facilitate a host of both scientific and business opportunities. Defra later made a commitment to publishing 8,000 data sets as 'open data' by June 2016 under the initiative #OpenDefra.

Specifically, in regard to bovine tuberculosis (bTB), an independent review of Defra's TB strategy [2] in England concluded that industry must take greater responsibility for on-farm controls, biosecurity and

safe trading. Government needs to supply more and sharper information on TB risks to facilitate that.

However, the wish for increased accessibility to data has to be weighed against the requirement to adhere to data protection considerations. The European Union (EU) General Data Protection Regulation (GDPR) is an EU Regulation and no longer applies to the UK since it left the EU. However, it has been retained in UK law (the Data Protection Act 2018 [3]) as UK GDPR. The aim of the UK GDPR is to give individuals (data subjects) more control over their personal data and its processing. It sets out a number of key requirements, which include requiring the consent of subjects for data sharing and processing and the anonymising of collected data to protect privacy. The implications of failing to comply with the UK GDPR can be serious, with fines in the millions of pounds having been handed out to companies and individuals who have failed to comply in recent years.

The data provided to researchers must conform to the seven key principles of UK GDPR. In particular the researchers must have a clear intended use for the data (purpose limitation) and they should only be provided with the data necessary to service that use (data minimisation).

These two drivers (the wish to both increase accessibility to data and adhere to the UK GDPR) have the potential to pull an organisation in opposite directions. Indeed, there is a danger that a strict interpretation of the UK GDPR by organisations (including Government) can stifle access to its data and the consequent benefits to be gained from any analysis.

Since the UK GDPR relates to the collection and processing of personal information, the first consideration when looking to share data is to decide whether the data to be shared include personal information. If they do, then the restrictions required by the UK GDPR apply. Personal data are defined as information that relates to an identified or identifiable individual. The data may be obviously and directly of a personal nature e.g. name and address or indirectly personal meaning the data could be used to identify an individual.

In the main, scientific researchers are not interested in personal information. In addition, research ethics committees will also place limitations around researchers accessing personal information, to ensure that any research participants are anonymised.

However, researchers may require data (e.g. a farm or animal identifier) that could potentially be linked to personal data i.e. is indirectly personal data. The decision as to whether such data are indirectly personal is often not straightforward and will need to be decided on a case-by-case basis.

In general, direct personal data are only required for projects such as socio-economic surveys or the identification of farms for on-farm studies. These types of studies are relatively rare and are far outnumbered by the requirement for data for epidemiological studies where direct personal and confidential data are not required.

Where data are deemed to be personal (either directly or indirectly), UK GDPR places restrictions on how they are shared e.g. requiring the prior consent of the data subject. In general, obtaining prior consent in the case of bTB would be unduly onerous e.g. Great Britain (GB) bTB testing records exist for more than 71,000 'live' herds. The UK GDPR does, however, allow the sharing of personal data where it is seen as being in the public interest and the use of personal data in research and development that helps to protect public and animal health is normally seen as compatible with the UK GDPR public interest exemption. This is the legal basis on which Animal and Plant Health Agency (APHA) shares key animal health data that are deemed personal.

In addition to the requirement to adhere to the UK GDPR, the sharing and publication of bTB data are also covered in the Tuberculosis in Animals (England) Order 2021 [4] – 'The TB Order'. The order sets out the rules and regulations relating to the control of this notifiable disease and, with regards to data sharing states:

Publication of information

23.— (1) The Secretary of State may publish information regarding how long a bovine herd has had officially tuberculosis-free status for the purpose of helping other persons to protect against the further spread of tuberculosis.

(2) Where a bovine herd loses its officially tuberculosis-free status the Secretary of State may publish information regarding that herd for the purpose of helping other persons to protect against the further spread of tuberculosis.

(3) Information which may be published under paragraph (1) or (2) may be published in any form that the Secretary of State sees fit.

As can be seen from the above, the TB Order is all-encompassing and allows the publication of data if it supports the public interest task of reducing the spread of infection.

Another important principle of the UK GDPR is 'data minimisation' – meaning that any data shared should only be sufficient for the data requesters needs. The practical implication of this is that the 'data owner' cannot simply hand over all data and leave the task of extracting what is required to the requestor but must take appropriate steps to ensure only the data required are shared.

Bovine tuberculosis data

Of all the diseases dealt with by APHA, bTB is associated with the most abundant and comprehensive data. It is a notifiable disease with a compulsory cattle testing regime in place. This testing, which is carried out by APHA (or on their behalf), has been electronically recorded since the mid-1980s meaning that an enormous amount of data has built up and is available for analysis.

Bovine TB can be considered a highly 'politicised' disease, with entrenched views as to its causes and how it should be managed. This makes the need for data sharing to be transparent and equitable very important, possibly more so than for any other animal health disease in GB.

Although the relational database holding this information consists of over 30 tables, from an epidemiological point of view the main data tables are:

- **Herd** – Contains data relating to the cattle herd being tested e.g. its location and type (beef, dairy etc.). Currently there are approximately 240,000 herd records. It should be noted that only approximately 71,000 of these are current 'live' herds in GB, the remainder of these records relate to 'archived herds' (herds that no longer exist).

- **Test** – Contains data relating to the testing of a herd (or sub-set of it) for bTB. Tests fall into two categories. Firstly, surveillance tests (the frequency of which is primarily determined by the prevalence of bTB in the location of the herd – currently either every six months, one year or four years) and secondly control tests carried out after bTB is identified or suspected. These tests must be negative before the restrictions associated with bTB occurrence are lifted. Currently there are approximately three million testing records from 1986 onward.

- **Breakdown** – Data relating to the bTB herd incident (or 'breakdown') that are recorded once bTB positive animals are found (or suspected) in a herd e.g. the start and end of the movement restrictions that are imposed on such herds. Currently there are approximately 98,000 breakdown records from 1986 onward.

- **Animal** – Details of all cattle tested e.g. sex and age. Up until 2012, only the details of animals that tested positive were recorded, but since that time details for all animals subject to a test have been recorded. For ease of processing these data are split into datasets with positive animals (currently approximately 1.5 million records from 1986 onward) and test negative ones (approximately 95 million from 2012 onward).

Although the GB bTB testing data are the main data set relating to bTB, it is only one of many. An example of another major bTB related dataset

is from the Randomised Badger Culling Trial (RBCT). The RBCT was a large study looking at the effect of badger culling on cattle bTB incidence in which the study treatments were carried out between 1999 and 2005. In common with other research or trial-based studies, these data were not made available for data sharing until the study had concluded, including the publication of relevant peer-reviewed papers. In general, after quality assurance through peer review, Defra and the devolved administrations are keen to maximise the value of such studies by making the data available for further analysis.

Although not specifically a bTB data-set, the Cattle Tracing System (CTS) data are often used in tandem with bTB data to analyse the effect of cattle movement on bTB incidence. Originally developed in response to the bovine spongiform encephalopathy crisis [5], the CTS contains the birth and death records of all cattle along with details of every movement to or from farms and markets and to slaughter. The CTS data therefore represents a remarkable store of GB cattle movement data – it currently consists of ~300 million individual movements of animals since the recording of cattle movements became compulsory in 2001. Although the primary reason for the recording of cattle movements is for regulatory and tracing purposes, the accumulated dataset represents an extremely valuable epidemiological and research asset.

Sharing data

Data requestors cover a wide spectrum both from within government and externally. Internal requests (from within Defra and its associated agencies) come from epidemiologist, statisticians and policy makers working in the field of bTB. External requests range from students requiring bTB data for undergraduate projects at one end to consortia of university research departments carrying out three-year research projects at the other. It should be noted, that in general, only data sharing external to the government department holding the data, is subject to the requirement for a data sharing agreement to be in place.

In general, external requests will require a data sharing agreement. For such requests APHA liaises with the requestor about their data requirements and creates a request specification that can be inserted into

a confidentiality agreement. If, however, the request is made under the Freedom of Information (FOI) act [6] the request must be dealt with exactly as set out in the request and there is no interaction with the requestor. Because of this, APHA would recommend that FOI's requests are not an ideal way of obtaining scientific research data.

Once a draft data sharing agreement has been created, approval is required from the Information Asset Owner (the person responsible for the information asset in the organisation that holds the data). This is likely to be more than one person if multiple data sets are requested. Permission to share data is also required from a person appointed by the national government if data from that nation has been requested.

Once all the necessary approvals have been given, the data are securely sent to the requestor along with appropriate meta-data. This usually involves uploading the files to a secure site and unique access details being sent separately to the requestor. It should be noted that the data agreement specifies that data access is time-limited and that after a specified date the data should be destroyed. Furthermore, access is given on a 'per project basis', meaning if access to the data is required for a different project, a new approval has to be given.

Over the past 20 years, the sharing of bTB data by APHA has enabled a wide-range of scientific research to be undertaken. This has ranged from traditional epidemiological and statistical analysis, spatial and social science analysis to more novel approaches such as machine, deep learning and genetic analysis. The work involved in the creation of bespoke datasets for analysis should not be underestimated, together with the provision of expertise relating to the available data it forms a vital part of any project. As such, those responsible for assembling the project data should be recognised in any peer-reviewed papers resulting from the work and the vast majority of lead authors are more than happy to include those responsible for data provision as co-authors.

The following examples illustrate the range of work that have utilised APHA's bTB data in recent years:

- The analysis of the RBCT data by Donnelly *et al.* [7] which aimed to establish the effectiveness of badger culling in reducing the incidence of bTB in cattle.

- The work of Gilbert *et al.* [8] which definitely established the link between cattle movement and bTB infection.

- Van Tonder *et al.* [9] have combined APHA-based genetic data with the testing data described above to better understand the relationship between badger and cattle infection.

- Examples of machine learning include the work of Banos *et al.* [10] to create an index (TB advantage) of resistance to TB infection in dairy cattle which is now used commercially.

- Denholm *et al.* [11] have used deep-learning techniques to try and predict bTB occurrence from the spectral analysis of milk samples using deep learning analogy.

The majority of analyses combine a number of data sets e.g. Donnelly *et al.* combined bTB testing data and RBCT data, Gilbert *et al.* combined bTB testing data and CTS. The ability to combine bTB data easily and accurately from different sources highlights the power of having data in a relational database format.

Publishing data

Defra has long published bTB breakdown numbers in an aggregate form as accredited National Statistics [12]. Such data provide a narrative to the general public on the progress in dealing with bTB at a national and county level and the bTB testing effort but are not designed to provide details relating to individual incidents.

Since the early 2000s there has been a drive from Defra to encourage farmers to take greater 'ownership' of the bTB problem and to support that objective they introduced a series of measures to encourage them to introduce better on-farm biosecurity, and to pursue 'informed' or 'knowledge-based' purchasing – i.e. to select animals for purchase taking the bTB history of the herd they came from into account. In

general, both the institutions representing the farming community (e.g. farming unions) and Defra agreed that, to enable farmers to take these steps, better access to bTB data was required. Defra accordingly committed to publishing more detailed bTB information at the level of individual herds. This represented a radical change in thinking, as up until that point farmers' bTB history had been deemed their personal information which should not be made public.

The first step in the process of publishing herd data was for Defra Ministers to introduce a new statutory power in the Tuberculosis (England) Order 2014 (since replaced by the Tuberculosis in Animals [England] Order 2021 referred to above). Once this had been achieved, it was decided that the most effective way to publish the herd level data would be in an interactive spatial format and the Environmental Research Group Oxford, in partnership with APHA, were commissioned to create 'information bovine TB' or 'ibTB' (https://ibtb.co.uk) to map the locations of bTB breakdowns. The initial version of the system first went online in the summer of 2015 and displayed current (live) incidents of bTB and historical ones occurring during the previous five years (later upgraded to ten years). Figure 1 shows ibTB at start-up displaying the current (live) incidents of bTB in England and Wales. From this initial view users can drill-down to their area of interest; as they do this the clusters of incidents will become smaller until (at the finest scale) only individual incidents are displayed.

Since its inception, a series of upgrades have been implemented – the latest of which went live on 31 January 2022 and allows users to create bespoke displays of farms that have been free of bTB for a given number of years. The number of years a herd has been bTB free is a good (and easy to communicate) proxy for bTB risk status [13]. This upgrade is only available for English herds because the ability to publish this information has not yet been approved by the Welsh Government. Scotland has Officially Tuberculosis Free status and their bTB incidents have never formed part of the system. ibTB is also linked to the commercial trading app SellMyLivestock (https://sellmylivestock.com) to encourage farmers to take account of the bTB risk in purchases they make.

In 2018 a series of usability trials carried out on farmers and veterinarians using the system rated its usability as 'excellent' [14].

Since its inception ibTB has been visited over 715,000 times by 183,000 unique users, showing there is an appetite from the farming community for access to this type of information and the system is now seen as a key component of Defra's advice for minimising bTB risk [15].

Because ibTB is a free access system that does not require users to register prior to use, it is not currently possible to know exactly who is using it. It is certainly being used by the farming community to better help farmers understand the disease picture in their area, as well as inform their purchasing decisions. But there have also been some unforeseen uses of the data e.g. it has been used by groups opposed to badger culling as a basis for their analyses.

Discussion

The sharing of animal data in an efficient manner is crucial both for encouraging and enabling scientific research and for successful disease control. However, it is important that the privacy rights of the data subjects (farmers) are adhered to by conforming to UK GDPR specifications, but equally such regulation should not become an impediment to data sharing where this is in the public interest.

Bovine TB in GB has a wealth of good quality data associated with it going back over 30 years and these data have underpinned much high quality and innovative research during that time. In more recent years it has been published in an accessible and user-friendly way via ibTB to empower farmers to take action to protect their herds from the disease.

The ibTB system has proved to be an effective way to proactively publish bTB data in a meaningful way. In general, the reaction to the system has been positive with only a few (under ten) farmers objecting to either Defra or APHA about their bTB history being made public. The system is of course limited by the availability and accuracy of bTB data. The locational data available is restricted to the location of the

affected farm and some farmers have made it known they wish to see the data at a finer scale e.g. the location of infected animals – data which are not currently collected.

The lack of detailed knowledge about the users of the ibTB system currently makes any meaningful evaluation of its impact impossible. For example, if the users of the system were known, it would be possible to understand what impact (if any) it had on their trading decision-making. The only way to understand the user base would be to implement a registration system. Such a system has been suggested but has met with some resistance from both policy-makers and users, mainly on the basis that it is seen as a potential impediment to use. Even if such a registration system was implemented in the future, it is highly likely that it would be voluntary, which would limit its usefulness from a user analysis point of view.

Despite the use that has been made of bTB data over the past 30 years, more can and should be done to encourage its use by making the research community and other interested parties more aware of what is available and making access to it easier and more efficient. The most obvious way of doing this would be the development of online access to meta-data describing the data and the ability to download those data that were non-personal and non-confidential. The creation of such an online portal for bTB data would be a significant piece of work but we believe the effort to create it would pay dividends in the professional and public interest benefits it would bring. Another option (which is currently in progress) is the development of an Application Programming Interface (API) allowing access to the data that underpins ibTB. The development of such an API will allow third party apps (e.g. cattle trading apps) to more readily utilise bTB data.

We believe the quality of animal health research and the ability to use the data to inform farmer's biosecurity decision-making is in direct correlation to the quality and availability of the relevant data. The analysis and management of other animal health diseases would benefit if the epidemiological data were more pro-actively shared, using systems like ibTB as a template.

Acknowledgements

This work has been funded by the bovine tuberculosis surveillance programme of Great Britain under contract no. SB4500.

References

[1] Department for Environment, Food and Rural Affairs (Defra) (2013). – Defra Open Data Strategy. Refreshed as part of the United Kingdom Open Government Partnership Action Plan. Defra, London, United Kingdom, 26 pp. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/267934/pb14109-defra-open-data-strategy-131219.pdf (accessed on 16 August 2022).

[2] Department for Environment, Food and Rural Affairs (Defra) (2014). – The strategy for achieving officially bovine tuberculosis free status for England. Defra, London, United Kingdom, 85 pp. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/300447/pb14088-bovine-tb-strategy-140328.pdf (accessed on 16 August 2022).

[3] Information Commissioner's Office (ICO) (2021). – Guide to the UK General Data Protection Regulation (UK GDPR). ICO, Wilmslow, United Kingdom. Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr (accessed on 16 August 2022).

[4] The Secretary of State (2021). – The Tuberculosis in Animals (England) Order 2021. Defra, London, United Kingdom, 20 pp. Available at: https://www.legislation.gov.uk/uksi/2021/1001/contents/made (accessed on 16 August 2022).

[5] Mitchell A., Bourn D., Mawdsley J., Wint W., Clifton-Hadley R. & Gilbert M. (2005). – Characteristics of cattle movements in Britain – an analysis of records from the Cattle Tracing System. *Anim. Sci.*, **80** (3), 265–273. https://doi.org/10.1079/ASC50020265

[6] Information Commissioner's Office (ICO) (2022). – How to access information from a public body. ICO, Wilmslow, United Kingdom. Available at: https://ico.org.uk/your-data-matters/official-information (accessed on 16 August 2022).

[7] Donnelly C.A., Woodroffe R. […] & Morrison W.I. (2006). – Positive and negative effects of widespread badger culling on tuberculosis in cattle. *Nature*, **439** (7078), 843–846. https://doi.org/10.1038/nature04454

[8] Gilbert M., Mitchell A., Bourn D., Mawdsley J., Clifton-Hadley R. & Wint W. (2005). – Cattle movements and bovine tuberculosis in Great Britain. *Nature*, **435** (7041), 491–496. https://doi.org/10.1038/nature03548

[9] Van Tonder A.J., Thornton M.J., Conlan A.J.K., Jolley K.A., Goolding L., Mitchell A.P., Dale J., Palkopoulou E., Hogarth P.J., Hewinson R.G., Wood J.L.N. & Parkhill J. (2021). – Inferring *Mycobacterium bovis* transmission between cattle and badgers using isolates from the Randomised Badger Culling Trial. *PLoS Pathog.*, **17** (11), e1010075. https://doi.org/10.1371/journal.ppat.1010075

[10] Banos G., Winters M., Mrode R., Mitchell A.P., Bishop S.C., Woolliams J.A. & Coffey M.P. (2017). – Genetic evaluation for bovine tuberculosis resistance in dairy cattle. *J. Dairy Sci.*, **100** (2), 1272–1281. https://doi.org/10.3168/jds.2016-11897

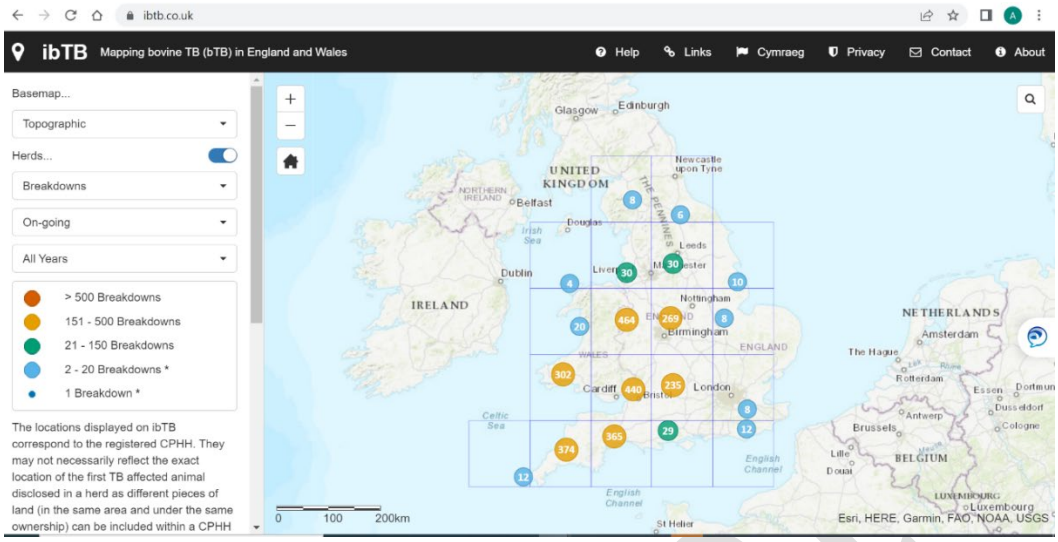
[11] Denholm S.J., Brand W., Mitchell A.P., Wells A.T., Krzyzelewski T., Smith S.L., Wall E. & Coffey M.P. (2020). – Predicting bovine tuberculosis status of dairy cows from mid-infrared spectral data of milk using deep learning. *J. Dairy Sci.*, **103** (10), 9355–9367. https://doi.org/10.3168/jds.2020-18328

[12] Department for Environment, Food and Rural Affairs (Defra) & Animal and Plant Health Agency (APHA) (2022). – Latest national statistics on tuberculosis (TB) in cattle in Great Britain – quarterly. Defra & APHA, Addlestone, United Kingdom. Available at: https://www.gov.uk/government/statistics/incidence-of-tuberculosis-tb-in-cattle-in-great-britain (accessed on 16 August 2022).

[13] Adkin A., Brouwer A., Simons R.R.L., Smith R.P., Arnold M.E., Broughan J., Kosmider R. & Downs S.H. (2016). – Development of risk-based trading farm scoring system to assist with the control of bovine tuberculosis in cattle in England and Wales. *Prev. Vet. Med.*, **123**, 32–38. https://doi.org/10.1016/j.prevetmed.2015.11.020

[14] Enticott G., Mitchell A., Wint W. & Tait N. (2018). – Mapping disease data: a usability test of an internet-based system of disease status disclosure. *Front. Vet. Sci.*, **4**, 230. https://doi.org/10.3389/fvets.2017.00230

[15] TB Hub (2020). – Reduce risk from neighbouring herds. Agriculture and Horticulture Development Board, Kenilworth, United Kingdom. Available at: https://ovp.6c6.myftpupload.com/preventing-tb-breakdowns/protect-your-herd-from-bovine-tb/reduce-risk-from-neighbouring-herds (accessed on 16 August 2022).



CPHH: county parish holding herd
ibTB: information bovine tuberculosis

Figure 1
The ibTB display at start-up