

RESEARCH

Open Access



# Systematic review of the status of veterinary epidemiological research in two species regarding the FAIR guiding principles

Anne Meyer<sup>1,2\*</sup>, Céline Faverjon<sup>1</sup>, Miel Hostens<sup>2</sup>, Arjan Stegeman<sup>2</sup> and Angus Cameron<sup>1</sup>

## Abstract

**Background:** The FAIR (Findable, Accessible, Interoperable, Reusable) principles were proposed in 2016 to set a path towards reusability of research datasets. In this systematic review, we assessed the FAIRness of datasets associated with peer-reviewed articles in veterinary epidemiology research published since 2017, specifically looking at salmonids and dairy cattle. We considered the differences in practices between molecular epidemiology, the branch of epidemiology using genetic sequences of pathogens and hosts to describe disease patterns, and non-molecular epidemiology.

**Results:** A total of 152 articles were included in the assessment. Consistent with previous assessments conducted in other disciplines, our results showed that most datasets used in non-molecular epidemiological studies were not available (i.e., neither findable nor accessible). Data availability was much higher for molecular epidemiology papers, in line with a strong repository base available to scientists in this discipline. The available data objects generally scored favourably for Findable, Accessible and Reusable indicators, but Interoperability was more problematic.

**Conclusions:** None of the datasets assessed in this study met all the requirements set by the FAIR principles. Interoperability, in particular, requires specific skills in data management which may not yet be broadly available in the epidemiology community. In the discussion, we present recommendations on how veterinary research could move towards greater reusability according to FAIR principles. Overall, although many initiatives to improve data access have been started in the research community, their impact on the availability of datasets underlying published articles remains unclear to date.

**Keywords:** Veterinary epidemiology, FAIR, Data access, Salmonids, Dairy cattle

## Background

The FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles were first published in 2016, providing a foundation to support increased re-use of scientific data [1]. Mons et al. [2] highlighted that FAIR is a

continuum, which they represented with six levels, from “re-useless data” to “FAIR data with open access and functionally linked”. The ultimate goal of this set of principles is for research objects (such as datasets) to be rendered reusable and increasingly re-used. A number of articles on the topic of FAIR principles, from the original group of authors and others have been published since 2017. Jacobsen et al. [3] described some of the challenges and opportunities for implementation of each of

\* Correspondence: [anne@ausvet.com.au](mailto:anne@ausvet.com.au)

<sup>1</sup>Ausvet Europe, 3 rue Camille Jordan, 69001 Lyon, France

<sup>2</sup>Department of Farm Animal Health, Utrecht University, 3512 JE Utrecht, the Netherlands



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the FAIR principles, while Thompson et al. [4] outlined the tools and technologies that are already available to support the adoption of FAIR data management, as well as the functionalities which are still lacking in that respect. Such resources enable scientific communities to identify existing solutions before considering developing their own.

There is however little visibility on the progress achieved by researchers in specific disciplines since the publication of the FAIR foundational paper. At the time of writing, very few published evaluations of FAIRness in publications and datasets were found in the literature. Van Reisen et al. [5] reviewed the implementation of the FAIR principles reported in 100 randomly selected academic journal articles citing the foundational FAIR paper. They point out that life sciences represent the vast majority of the implementation, with 95 of the selected papers related to this discipline and a very limited representation (5 papers) of the other disciplines such as social science, humanities and other sciences. As supplementary materials to their publication on FAIR metric development, Wilkinson et al. [6] also provided evaluation results for ten digital resources, as example application of these metrics. In addition, two reviews on application of FAIR principles, in Europe and Africa respectively, have been published recently, gathering large scale initiatives which are representing important steps to optimise data management and stewardship, and therefore, strive towards more FAIRness [7, 8]. These reviews highlighted initiatives in humanities, environmental science, materials science and digital health, but none in many other sectors, such as veterinary research.

The importance of veterinary research in general, and especially veterinary epidemiology, has been better acknowledged since the beginning of the COVID-19 pandemic. Managing and improving animal health is not only critical to provide a secure access to safe animal-sourced foods, but also as a key component of the global health ecosystem. The purpose of the work presented in this manuscript is to fill the current gap of knowledge about the adoption of FAIR principles in this discipline. Our study is based on a systematic review of the FAIRness of datasets associated with peer-reviewed articles relating to veterinary epidemiology research and published since 2017. The objectives are (i) to assess the state of FAIRness in this discipline, currently and over the past few years, and (ii) to explore how veterinary epidemiology research could move towards more reusability, in line with the goal of the FAIR principles.

Many tools for evaluating the FAIRness of digital resources are available, in the form of questionnaires. Some of these tools were reviewed in peer-reviewed publications [4, 9], while others are referenced in online repositories such as the FAIRassist [10] and Research

Data Alliance [11] repositories. Wilkinson et al. [6] in particular have proposed a framework and a first set of metrics developed for the evaluation of FAIRness. The authors subsequently proposed a second set of metrics, called maturity indicators, after including community feedback regarding the first set and gaining a better understanding on how data generators, managers and users were addressing FAIR principles [12]. These maturity indicators are registered by *FAIRsharing* ([www.fairsharing.org](http://www.fairsharing.org)), an online register of metadata standards to allow scientists to use frameworks which have been thoroughly documented. In the present manuscript, we applied the maturity indicators and tools proposed by these authors to conduct our evaluations. This framework intends to evaluate the overall maturity of an approach by assessing the FAIR maturity indicators separately and identifying specific points that can be improved, rather than evaluating the resource with a summary score of FAIRness: "FAIRness is not a competition, rather, FAIRness refers to a maturation process where digital objects are rendered increasingly self-descriptive to the machine" [12].

Given the diversity found in the animal production sector in terms of species, further definition of the scope of this study is required. Aquaculture is an increasingly important provider of animal protein worldwide [13]. There are dozens of diseases which affect the economic sustainability of aquaculture enterprises [14] and may be of concern regarding public health [15]. In this context, epidemiological studies provide key tools to better understand the complex systems in which fish and other aquatic species are produced [16], by looking at the variations in disease risk in populations and considering the interplay between host, pathogen and environment factors. Epidemiological research to improve aquatic health management, and salmonid health in particular, emerged as a discipline in the early 2000s and is still growing [17, 18]. Salmonids (including several salmon and trout species) are an important group in seawater fish production globally, with over 3 million tons produced in 2016 [13], and also an interesting group for the present work given the increasingly common use of data routinely generated by commercial producers for research. Such data are likely to be considered as confidential information, due to animal health and production data revealing production practices, adding another layer of complexity in terms of FAIRness. Given the growing importance of this sector, this study focuses on the current state of FAIRness in salmonid epidemiology.

In contrast, milk and dairy products are contributing to a much larger proportion of food produced from animals worldwide than fish [19], making dairy cattle one of the major livestock production systems. However, there are increasing concerns about the impact of terrestrial livestock production systems on the environment

and on climate change. It is estimated that 15% of human-induced emissions of greenhouse gases worldwide are attributable to livestock production, with a large proportion of these emissions due to ruminants [20]. Nonetheless, ruminants, and dairy cattle in particular, remain critical in many ecosystems given their unique ability to convert feedstuffs with little nutritional value for humans into high-quality protein [21]. In this context, epidemiological research in dairy cattle is a long-established discipline, as improving animal health is key to optimise yields and thus make the most of available resources. Given the importance of dairy cattle in epidemiological research, we chose to compare our snapshot of the state of FAIRness in salmonid epidemiological research with a similar snapshot in this species to address the two study objectives stated above.

In this work, molecular epidemiology research, the branch of epidemiology using genetic sequences of pathogens or hosts to describe disease patterns, was considered separately from other branches of epidemiology, given the differences in the types of data that are collected and analysed in these disciplines. The results presented in this manuscript concern both molecular and non-molecular epidemiology in salmon, and non-molecular epidemiology only in dairy cattle.

## Results

### Data availability in salmonid research

The literature search yielded 147 and 98 results for salmon and trout, respectively. The flow diagram of the identification, screening and inclusion of results is included in Supplementary File 1. The review process led

to the inclusion of 91 articles on various epidemiological topics in salmonid production in this study (the full reference list is provided in Supplementary File 2). The selected articles presented data from 16 individual countries, with an important representation from Norway, Chile and Canada, as well as global data and grouped data from North America, Latin America and Europe (Fig. 1, left panel).

The assessment of the full texts, supplementary materials and article pages on the publisher websites showed that for 55 out of the 91 papers, the raw data supporting the work were not provided within the article, via its supplementary materials or in an online repository (Fig. 2). Raw data were available for 80% of the molecular epidemiology papers (24 out of 30) and 20% of the papers in other epidemiology sub-disciplines (12 out of 61). In addition, authors stated that raw data were available upon request in 3 out of the 55 papers for which they were not directly provided.

For the 24 molecular epidemiology papers with raw data classified as available, genetic data were uploaded in specific-purpose repositories: GenBank (19 papers), NCBI's Sequence Read Archive (4 papers), PubMLST (3 papers) and the European Nucleotide Archive (2 papers), or in a generic-purpose repository (FigShare, 2 papers). The sum of the numbers in brackets is larger than 24 as six papers deposited data in two repositories. Genetic data were not available for the six remaining papers. Epidemiological data on the isolates or samples were also provided in two thirds of molecular epidemiology papers (21 out of 30). Such data were generally shared as tables within the manuscript itself (14 papers), and/or as supplementary

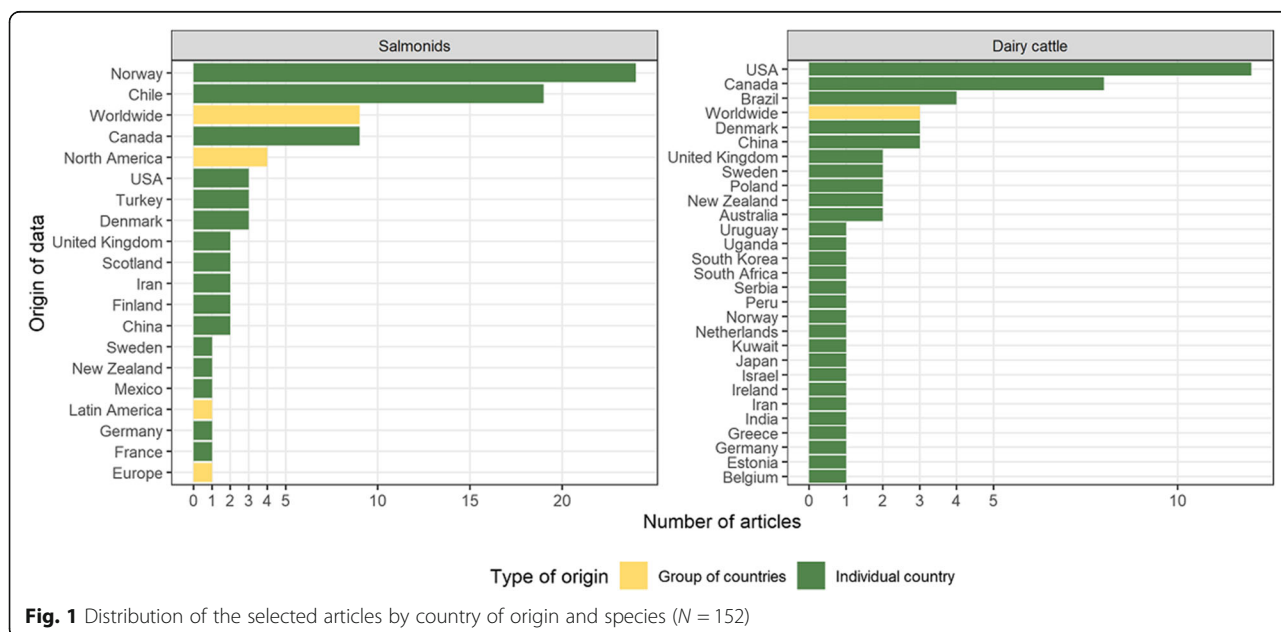
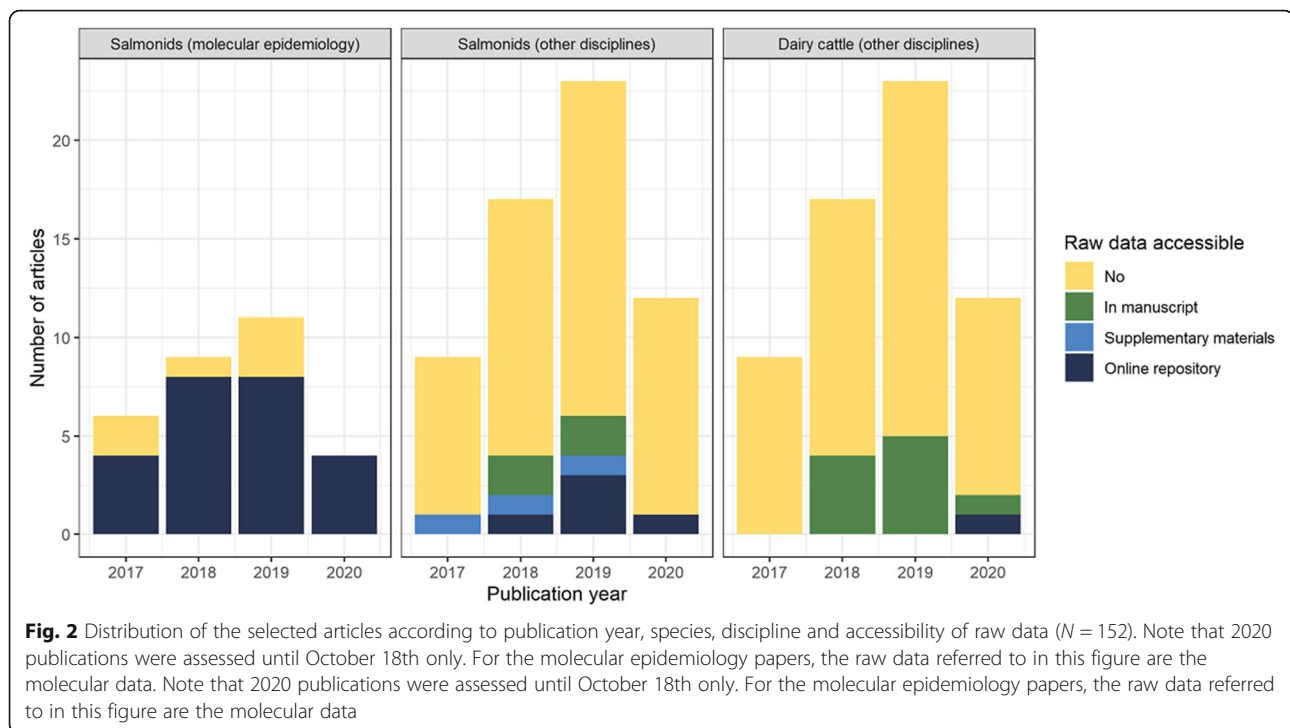


Fig. 1 Distribution of the selected articles by country of origin and species (N = 152)



materials (either as PDF tables, 3 papers, or Word tables, 5 papers). Isolate data were shared in a MicroReact project for one paper and could be downloaded as a tab-separated file.

For the 12 papers in other epidemiology sub-disciplines for which raw data were available, these were deposited in an online repository in 5 cases: a generic-purpose repository (Mendeley Data, Dryad) or institutional repositories (Marine Data BC and Norwegian Marine Data Centre). In the other cases, the raw data were shared in the manuscript itself as Tables (4 papers), as supplementary materials (either as PDF tables, 1 paper, or Excel file, 2 papers). For two of these papers, it appeared that only part of the data used to conduct the work was made available. Last, ten of the 61 non-molecular epidemiology papers used commercial data, but raw data were available for only one of these ten papers.

#### Data availability in dairy cattle research

At the end of the eligibility and inclusion steps (see flow chart in Supplementary File 1), a total of 61 papers concerning epidemiology research in dairy production were randomly selected and assessed for data availability. A third of the papers related to Canada or USA (20 papers out of 61), while the remainder originated from 26 other countries or were based on worldwide data (3 papers) (Fig. 1, right panel). The assessment of dairy papers showed that raw data were available for 18% of them (11 out of 61) (Fig. 2). In ten of those, the raw data were

provided within the manuscript. Raw data were deposited in a generic-purpose repository for the remaining paper (Scholars Portal Dataverse). In addition, authors stated that raw data were available upon request in 3 out of the 50 papers for which they were not directly provided. Four of the 61 dairy cattle papers used commercial data, but raw data were not available for any of these four papers. An overview of the data availability in both species is presented in Table 1.

#### Data availability statements

The 152 articles assessed in this part of the study were published in 60 different journals. Among those, 17 included a formal data availability statement either as a dedicated section in the manuscript or as a supplementary item (Fig. 3). Such statements were introduced during the study period in some of the journals most represented by the selected articles, such as Aquaculture, Journal of Fish Diseases and Preventive Veterinary Medicine. These additional sections may be named “Availability of data and materials”, “Data access”, “Data accessibility”, “Data availability”, “Data profile”, “Data summary” or “Research data for this article”. Among papers not related to molecular epidemiology, data availability statements were provided in 6 and 10 of the dairy and salmonid articles, respectively. In salmonid papers, the statements mentioned that the study datasets were available upon request (2 papers), that authors did not have permission to share them (1 paper), that the data had been deposited in an online repository (4 papers), or that “all relevant data are within

**Table 1** Overview of data availability in articles included in this review

Species (discipline)	Salmonids (molecular epidemiology)	Salmonids (other disciplines)	Dairy cattle (other disciplines)
Number of articles included in the review	30	61	61
Number of articles which had raw data available	24	12	11
• In manuscript	0	4	10
• As supplementary materials	0	3	0
• In an online repository	24	5	1
Number of datasets included in the FAIRness assessment	30	8	1

the paper and its Supporting Information files” (or similar, 3 papers). In two of the three latter papers, the raw data did not appear to be available despite the statement. In dairy papers, the statements mentioned that study datasets were available upon request (3 papers), that authors did not have permission to share them (2 papers), or that the data had been deposited in an online repository (1 paper).

Some individual journals may appear in more than 1 year, species or discipline. Note that 2020 publications were assessed until October 18th only.

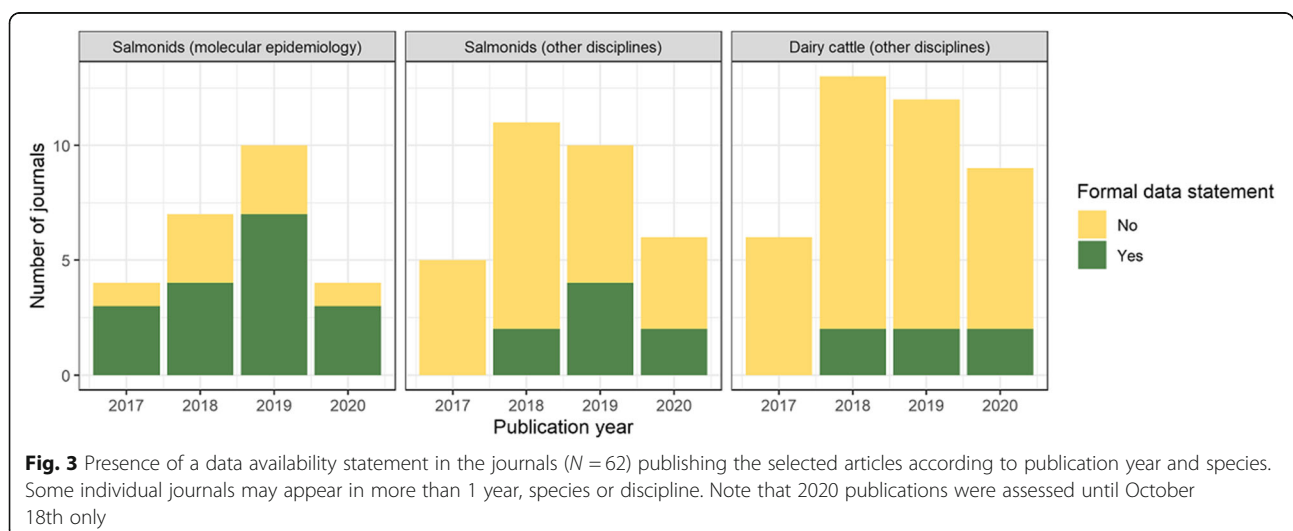
#### FAIRness assessment

Evaluation results for 13 maturity indicators of the available datasets are presented in Table 2 (papers on molecular epidemiology) and Table 3 (other papers). The assessment criteria are described in the Methods section (Table 4). A number of papers provided the data used in the study within the manuscript itself (text or tables) (Fig. 2), including 4 and 10 papers on salmonids and dairy cattle, respectively. Those datasets were not evaluated for FAIRness as they do not form a distinct digital resource from the article itself.

Molecular digital resources were almost always identified by a globally unique identifier defined by the repository, but this identifier was generally not persistent (F1).

Most resources were associated with metadata including the data identifier (F2 and F3) and provenance information (R1.2). All resources were indexed by a search engine (F4) and accessible via an open, free protocol (A1) but the persistence of the metadata should the resource become unavailable was not guaranteed (A2). No resources used a formal knowledge representation language, but all were provided in standardized formats, such as GenBank data (I1). Most used FAIR vocabularies (I2) and about half were linked to other relevant resources (I3). Clear license conditions were not often provided (R1.1), given that several common repositories such as GenBank state that the repository managers “cannot provide comment or unrestricted permission concerning the use of the information contained in the molecular databases” or similar. All repositories but one were certified or listed as trusted in known community schemes (R1.3). The digital resources evaluated here originated from a range of 12 individual countries and from two multi-country studies.

Within non-molecular datasets, most digital resources were identified by a globally unique and persistent identifier, generally a Digital Object Identifier (F1). Six out of nine resources were associated with metadata including the data identifier (F2 and F3) while most had



**Table 2** Evaluation of FAIRness for 30 datasets accompanying 24 articles in molecular epidemiology of salmonids

Maturity indicator assessment													Source article
F1.1	F1.2	F2	F3	F4	A1	A2	I1	I2	I3	R1.1	R1.2	R1.3	
0	0	2	0	2	2	0	1	2	2	2	2	0	[22] (PubMLST) + [23] (PubMLST) + [24] (PubMLST)
2	0	2	2	2	2	2	1	2	2	1	2	2	[22] (European Nucleotide Archive) + [25] (European Nucleotide Archive)
2	0	2	2	2	2	0	1	2	0	1	0	2	[26, 27] + [28] (Sequence Read Archive)
2	0	2	2	2	2	0	1	2	0	1	2	2	[29–35] + [23] (GenBank) + [24] (GenBank) + [28] (GenBank)
2	0	2	2	2	2	0	1	2	2	1	2	2	[36–43] + [44] (GenBank)
2	2	1	2	2	2	2	1	0	0	2	0	2	[45]
2	2	1	2	2	2	2	1	2	0	2	0	2	[25] (FigShare)
NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	[44] (Sequence Read Archive)

The scores correspond to whether a given maturity indicator was met (2), partially met (1) and not met (0), respectively (cf. Table 4). The mention NE corresponds to a resource which could not be evaluated given that the dataset accession number was invalid. Datasets which have obtained the same combination of scores on the 13 indicators are grouped in the same row. Given that six source articles had two datasets associated with them (deposited in separate repositories), their reference appears twice in this table. To distinguish them, the corresponding repository is indicated in brackets after the reference.

provenance information (R1.2) and were indexed in a search engine (F4). Accessibility indicators were similar to those observed above for molecular data (A1, A2). No resources used a formal knowledge representation language and only two were provided in editable, non-proprietary format (I1). No resources used standardized vocabularies, and only two provided some documentation regarding the vocabularies used (I2). None were linked to other relevant resources (I3). Clear license conditions were always provided (R1.1) and five datasets were deposited in repositories certified or listed as trusted in known community schemes (R1.3). The nine digital resources which were evaluated originated from a range of countries: Canada (two datasets), Finland (one dataset), Mexico (one dataset), Norway (two datasets) and USA (one dataset) or from multi-country studies (two datasets).

Finally, the data sources for the nine non-molecular epidemiological digital resources evaluated in this section were databases from government agencies (e.g., the

Directorate of Fisheries in Norway [50] and the Aquatic Health Committee of Oaxaca in Mexico [49]) or international agencies (e.g., European Community Reference Laboratory for Fish Diseases [48]), data from the published literature or data specifically generated for the purpose of the study. The dataset made available by Soler-Jiménez et al. [54] in their literature review was the only dataset from a commercial source. However, this dataset was not collected specifically for the purpose of the study described in this publication. The authors were able to obtain mortality, environmental and management data from a group of fish producers in Mexico in the context of another study, which could not be identified in peer-reviewed sources at the time of writing. The dataset contained a few hundreds of farm-level records of several variables, without associated farm identifiers.

**Discussion**

Most datasets used in non-molecular epidemiological studies were not findable, or “re-useless” as characterised

**Table 3** Evaluation of FAIRness for nine datasets accompanying articles in non-molecular epidemiology papers

Maturity indicator assessment													Source article
F1.1	F1.2	F2	F3	F4	A1	A2	I1	I2	I3	R1.1	R1.2	R1.3	
2	2	0	0	2	2	0	0	0	0	2	0	0	[46]
2	2	2	2	2	2	2	1	1	0	2	2	2	[47]*
0	0	0	0	0	2	0	0	0	0	2	2	0	[48]
2	2	2	2	2	2	2	0	0	0	2	2	2	[49]
2	2	2	2	2	2	0	0	1	0	2	2	2	[50]
2	0	2	2	2	2	0	0	1	0	2	2	0	[51]
2	2	2	2	2	2	2	0	0	0	2	2	2	[52]
2	2	2	2	2	2	0	1	0	0	2	2	2	[53]
0	0	0	0	0	2	0	0	0	0	2	0	0	[54]

The scores correspond to whether a given maturity indicator was met (2), partially met (1) and not met (0), respectively (cf. Table 4). The mention NE corresponds to a resource which could not be evaluated given that the dataset accession number was invalid. Eight papers concerned salmonids and one dairy cattle (the latter is marked with an asterisk).

by Mons et al. [2]. These authors estimated that 80% of datasets in science belonged to this category. Our assessment suggests that the proportion in veterinary epidemiology is at least as high for the two species included in this work. In addition, no clear trend of improvement was observed over the past 4 years. Data availability was much higher for molecular epidemiology papers, with 80% of the articles assessed depositing genetic data in online repositories. The low proportion of raw data provided as distinct digital resources in non-molecular epidemiology publications means that only few datasets could be assessed in terms of FAIR maturity indicators in this study (nine out of 122 articles assessed). Of note, our assessment considered whether all the raw data used to produce the results were available for a given paper, but we did not try to reproduce any of the results as this was out of scope of the present study. Such additional assessment may allow the identification of papers for which some of the data required to reproduce the results were not available.

For articles with no raw data or, more rarely, data shared within the manuscript itself, the absence of separate data objects means that the FAIR assessment could not be conducted. Although these data may be reusable as they are both findable and accessible by researchers, and associated with provenance information and other metadata, their format (not directly searchable and editable) does not make them interoperable. In addition, they are neither machine-findable nor machine-accessible. Such considerations also apply to the data shared in supplementary files assessed in this study. For supplementary materials, metadata may not be readily available (F2), although researchers may manually find relevant information in the article, and these resources sometimes do not have a unique resource identifier (F1), although some journals associated a dedicated DOI to each supplementary item. Supplementary materials were not searchable in search engines, as the article itself is the object that is indexed (F4). Although providing raw data as supplementary materials does not make datasets FAIR, it is a common and practical way for researchers to respond to the increasingly pressing requests to make data available. It could be argued that there is little difference between sharing raw data in a manuscript table or in a supplementary material table and that treating them differently is arbitrary. However, a criterion for what constitutes raw data had to be established for the specific purpose of this study. In accordance with the FAIR framework which focuses on digital objects, we chose to only assess the FAIRness of raw data which were provided as separate digital resources, as described in the Methods section.

While the number of papers for which the raw data were available was similar between salmonid and dairy cattle in non-molecular epidemiology (12 and 11, out of 61, respectively), datasets were made available as individual digital resources more often by salmonid researchers than by dairy researchers (8 and 1 datasets, respectively). It was not possible to compare the FAIR maturity indicators between the two species given the small sample size and they are discussed here together. Across disciplines and without considering the case of supplementary material datasets which was already discussed above, the FAIRness assessment showed that most resources were uniquely identified, although persistence of these identifiers was generally not provided for molecular datasets (F1). Indicators F2, F3 and R1.2 were often met, with some metadata provided for the digital resource, along with the data identifier and provenance information. The persistence of the metadata should the data objects be removed was generally not guaranteed regardless of the repository considered. Importantly, all data objects were discoverable by web-based search engines (F4) and freely accessible via an open protocol (A1).

While the available objects generally scored favourably for Findable, Accessible and Reusable indicators, Interoperability was more problematic. The datasets identified in our study were not using a “formal, accessible, shared, and broadly applicable language for knowledge representation” as per the definition of the I1 indicator [1]. In non-molecular epidemiology, datasets were often shared as Excel files, with little to no documentation of the content of the different data fields, and no linked objects could be identified. Much progress remains possible on the I1, I2 and I3 indicators for these resources. Examples of existing resources which may be used to improve Interoperability in epidemiological research are the *AGROVOC* vocabulary developed and managed by the Food and Agriculture Organization (<http://www.fao.org/agrovoc/>) and the *SNOVET* systematized nomenclature for veterinary medicine [55]. By contrast, standardized formats (e.g., XML) and vocabularies are used in the National Centre for Biotechnology Information (NCBI)’s databases considered in this study for molecular datasets (GenBank and Sequence Read Archive) (23 out of 30 datasets). As the I1 maturity indicator is not associated with an agreed list of what constitutes a valid language for knowledge representation for a given discipline, there is room for interpretation by the person performing the assessment in terms of defining their scoring system. As such, the scoring system used in the present study may not be valid for another study or discipline. It is important to recognize that, as the coding of molecular data is universal, it is much easier to achieve interoperability for such datasets, as shown by

the good scores achieved on this indicator compared to non-molecular epidemiology. In addition, the digital archiving of such data is concentrated by a few stakeholders as NCBI is collaborating with other large repositories such as DNA DataBank of Japan and the European Nucleotide Archive, making it practical for the interested researcher to find related data, even if they are not explicitly linked (I3). Molecular epidemiology papers were also frequently providing epidemiological data (isolate or sample data) in a tabular format (70% of papers). Such data, which can be considered as raw data or as metadata for the genetic data, are critical to support reusability. Although inferior to the availability of genetic data (80%), the availability of epidemiological data was substantially higher than that observed in non-molecular papers. Finally, frequent re-use of published genetic sequences by other researchers is observed in the literature, showing the progress made in this discipline compared with other disciplines. In summary, the relatively consistent structure of molecular datasets, the availability of appropriate repositories, and the existing demand for data re-use are some of the factors which may explain the differences in data availability observed between molecular and non-molecular papers in our study.

Meeting indicators related to the Findable, Accessible and Reusable principles is likely possible for researchers who do not have specific skills in data management. On the other hand, interoperability appears to be a more complex objective. The process of migrating the Pathogen-Host Interaction Database, in plant sciences, to a FAIR-compliant form [56] illustrates that the data transformation required to apply machine-readable standards for knowledge representation require specialist knowledge in this area. This may be an obstacle for both funders and scientists who may not readily understand the concepts or have the skills required for effective data preparation, management and long-term preservation [57–59]. In this regard, more systematic data management training is needed in graduate programmes, both to develop awareness around open science and FAIRness and to teach specific skills required to reach these goals. Meanwhile, researchers may already take simple measures to increase the interoperability of their datasets. For example, documenting the content of tabular data in a systematic manner, indicating the content, type and unit of each data field, is an accessible step for rendering datasets self-descriptive, short of using knowledge representation languages and FAIR vocabularies. Units and conditions of measurement are critical for the re-use of quantitative data. In the case of ‘oxygen’ levels in fish cages, one needs to know the type and unit of the measurement (e.g., oxygen

saturation in %, oxygen concentration in mg/L) as well as the depth and temperature of the measurement to allow for correct interpretation.

No comparable assessments of FAIRness in publications were found in the literature but the subjects of data sharing and open data in research have been abundantly documented. A strong contrast is observed between willingness to share research data expressed by scientists and availability of datasets in practice. Survey respondents often declare willingness to share at least some data publicly (around 80% of them depending on the discipline and study) [60, 61]. Recent studies looking at the proportion of published work for which datasets were publicly available showed relatively low levels of availability: 8% in geoscience flux research [62], less than 10% in psychology research [57], 14% in morphology research [63] and 18% in biomedical research [64]. Even in biomedical journals with a full data sharing policy for randomized controlled trials, a review found that only 17 out of 37 eligible articles satisfied the definition for data availability [65]. Our present findings are consistent with these observations. Despite the plethora of published literature on the topic of data sharing and the growing availability of technological solutions, the evolution of practices and attitudes remains slow. In this regard, as our study was conducted only 4 years after the principles were formally published, it is not surprising that no significant changes were observed yet. These results will provide a baseline measure for future evaluations aiming at identifying longer-term changes.

Regardless, some scientific disciplines are significantly ahead in terms of data sharing, for example in genomics [66], as confirmed in the present study for molecular epidemiology in animal species. In other disciplines, barriers to making research datasets available have been explored [57, 61, 67, 68] and some of the factors which can positively impact data sharing have been identified [69–71]. Top-down pressure, in the form of strong encouragement or policies of mandatory sharing, from funding bodies and journal publishers appears to be highly effective. For measurable progress, such policies and requirements must not remain theoretical but need to be verified in practice. Data management plans indicating how FAIR principles will be applied are an increasingly common requirement from research funding organisations [72, 73].

In the animal production sector in particular, data provided for research by stakeholders may be considered as confidential information, due to animal health and production data revealing production practices as well as representing commercial assets and competitive advantages. Fourteen of the 122 non-molecular epidemiology papers reviewed in this study were based on commercial data, including four in dairy cattle and ten in salmonids.



The availability of raw data was very low for these potentially sensitive commercial production data (only one of the fourteen papers made the raw data publicly available). It is reasonable to assume that raw data provided by commercial producers are even more concerned by privacy issues than datasets collected by other means. In such cases, researchers are subjected to conflicting needs, with transparency, reproducibility and reusability on one side, and data confidentiality on the other side. Our results suggest that researchers may have more room to make datasets available when they collected these data from third-party sources, such as government databases holding data submitted by industry for regulatory purposes (for example the study by Myksovoll et al. [50]). Sourcing research data from third-party data integration initiatives may allow the generation of epidemiological datasets which are easier to share publicly, due to pre-existing data sharing agreements with the industry. An example of such initiative found during this review is *Fish-iTrends*, a sea lice data management system administered by the Atlantic Veterinary College in Canada [74]. Another example in salmonids is the attempt to set up a data integration platform described by Meyer et al. [75]. The use of animal production commercial data for research can significantly increase access to quality data with excellent coverage in time and space of the populations of interest, as shown by studies included in this work such as analyses conducted in the Chile salmonid industry [76, 77] and in the Canadian dairy industry [78]. The confidentiality level required for these datasets as well as barriers to data sharing related to competitiveness and anti-trust regulation aspects suggest that there could be a fundamental incompatibility between the principles of open-access data and the use of commercial data for research. However, progress towards more FAIRness does not require the datasets to be made open access. The four first levels presented by Mons et al. [2] are compatible with access-restricted datasets. Data integration systems such as those proposed above may allow progress to be made towards improved FAIRness. For example, authors making datasets findable and their metadata findable, accessible, interoperable and reusable would reach the fourth of these levels, “FAIR data with restricted access”. This would better enable other researchers to assess which digital resources could potentially be useful to them and express their interest in reusing such datasets. Negotiations regarding the conditions for access, such as the appropriate type of license, could then be undertaken between interested parties. This would represent significant progress from the “re-useless data” stage where most of the datasets assessed in the present study were found.

Finally, our assessment showed that molecular datasets appear relatively compliant with machine readability, while

this was generally not the case for the nine non-molecular available datasets. Thus, machine readability remains uncommon for non-molecular epidemiological data, especially as these nine datasets only made up for a small proportion of the 61 studies initially identified. The approach proposed by Wilkinson et al. [12] puts emphasis on machine readability: “detecting and validating behaviours of digital objects that make them machine-readable and reusable”. Given the relative novelty of FAIRness implementation in veterinary epidemiology, it is not surprising that only few resources were standardized for machine readability, while most provide human-readable content. In addition, the FAIRness assessments strictly focus on the digital object itself, and therefore do not assess other aspects of compliance with generic data management good practices, such as data curation and governance. The salmonid data platforms used in some of the studies [74, 76, 77] show the progress made by producers and data users to explore new ways of managing data, which require data curation and governance aspects to be defined in a collaborative manner. The FAIR principles also do not consider discipline-specific attributes which may be considered critical. For instance, some authors proposed an extension of the FAIR principles to better address the reproducibility and privacy protection challenges encountered in health research [79]. There are no standard indicators available to assess the additional aspects, and they were not considered within the scope of this study. Last, FAIRness assessments may also be applied to model code objects, when these are shared by their authors. Although we did not specifically address this aspect in the present work, we noted that none of the fourteen mathematical modelling studies provided access to such model objects.

## Conclusion

In conclusion, we recommend that the FAIR framework is progressively integrated in the routine workflow of researchers in veterinary epidemiology, starting with more education, training and communication. Furthermore, the barriers to reach the goals of data re-usability which have been set for a few years should be identified by discipline. This would allow the design and implementation of interventions to overcome these barriers. Last, we suggest a stepwise approach to improving the FAIRness of research data, in which the first step would be to make a large proportion of datasets and their metadata findable as digital resources.

## Methods

The methods used for this systematic review are reported in accordance with the PRISMA statement [80]. A literature search was conducted on October 18th, 2020 to identify all peer-reviewed publications related to epidemiological research in salmonid

**Table 4** List of 13 FAIRness maturity indicators evaluated in this study, based on the framework proposed by Wilkinson et al. [6, 12]

			<b>Maturity levels</b>
Indicator identifier	Indicator name	Indicator description	
F1.1	Identifier uniqueness	Whether there is a scheme to uniquely identify the digital resource	0: Indicator is not met 2: Indicator is met
F1.2	Identifier persistence	Whether there is a policy or scheme which ensures the persistence of the digital resource identifier	0: Indicator is not met 2: Indicator is met
F2	Data are described with metadata	Whether metadata corresponding to the digital resource are available	0: Indicator is not met 2: Indicator is met
F3	Resource identifier in metadata	Whether the metadata contains the unique identifier for the digital resource	0: Indicator is not met 2: Indicator is met
F4	Indexed in a searchable resource	Whether the digital resource can be found by web-based search engines using search terms such as title, author or key words. Google Search was used in this assessment	0: Indicator is not met 2: Indicator is met
A1	Access protocol	Whether there is an open and free access protocol to retrieve the digital resource, and if not, whether the specifications to access restricted content are provided	0: Indicator is not met 2: Indicator is met
A2	Metadata longevity	Whether there is a policy to guarantee the persistence of metadata even in the case of absence or removal of the digital resource itself	0: Indicator is not met 2: Indicator is met
I1	Use a knowledge representation language	Whether a formal language for knowledge representation is used in the digital resource. This indicator was assessed in terms of the format of the data.	0: Indicator is not met 1: Indicator is partially met when the resource is in a structured, non-proprietary, editable format (e.g., CSV, XML) 2: Indicator is met when the language used is cited and documented
I2	Use of FAIR vocabularies	Whether the digital resource uses formal and shared vocabularies (ontologies) for knowledge representation, which are themselves terms from open, community-accepted vocabularies published in an appropriate knowledge-exchange format.	0: Indicator is not met 1: Indicator is partially met when the vocabularies used in the resource are documented 2: Indicator is met
I3	Use of qualified references	Whether the digital resource or its metadata contain relationships with third-party data, with an explicit and useful semantic meaning	0: Indicator is not met 2: Indicator is met
R1.1	Accessible usage license	Whether there is a license document for the digital resource and the ability to retrieve those documents	0: Indicator is not met 1: Indicator is partially met when elements concerning the conditions for re-use, copying or distributing the resource are available but no formal license can be found 2: Indicator is met
R1.2	Detailed provenance	Whether the digital resource content is associated with provenance information associated with the data, covering at least: (i) who produced the data and when, and (ii) why and how the data was produced (context and relevance of the data). The availability of such information in the digital resource itself or its metadata was evaluated, not in the content of the associated article	0: Indicator is not met 2: Indicator is met
R1.3	Meet community standards	Whether the digital resource is listed by a recognized body as meeting community standards. Repositories were considered as compliant when listed by FAIRsharing, the Registry of Research Data Repositories or Core Trust Seal ( <a href="http://www.coretrustseal.org">www.coretrustseal.org</a> )	0: Indicator is not met 2: Indicator is met

The indicator identifiers are the same as the corresponding guiding principle identifiers for simplicity

production and in dairy cattle published between January 1st, 2017 to October 18th, 2020. We searched scientific literature referenced in three electronic databases: Scopus, Web of Science and PubMed. The documents retrieved were imported into a desktop reference management system for screening, eligibility assessment and further analysis.

For the salmonid dataset, we used the following keywords: (epidemiology) AND (salmon OR trout) AND (aquaculture). All titles, abstracts and key words were screened to select the results relevant to our study, i.e., articles related to epidemiological research in salmonid production. All farmed salmonid species were included (e.g., Atlantic salmon, coho salmon, rainbow trout). We excluded (i) studies in other disciplines (clinical reports, molecular biology, proteomics, bacteriology, microbiology, parasitology,<sup>1</sup> phylogenetics, physiology, food safety, economics, welfare), (ii) studies not conducted in farmed salmonids (wild salmonids, other species) and (iii) documents other than peer-reviewed publications (e.g., conference papers). Three additional articles were removed at the full-text review stage (see below) as they were discussion papers and therefore not based on any formal dataset.

For the dairy cattle datasets, we used the following keywords: (epidemiology) AND (dairy) AND (cattle OR cow). Given the large number of database results, only titles and key words were screened to select the results relevant to our study, i.e., articles related to epidemiological research in dairy cattle. Abstracts were reviewed only when the title and key words did not provide sufficient information. Exclusion criteria (i) and (iii) mentioned above were used. For criterion (ii), studies not conducted in dairy cattle (e.g., dairy goats) were excluded. Additional exclusion criteria were (iv) articles which were not available in English and (v) molecular epidemiology papers. The latter were not considered for inclusion, as the comparison between salmonids and dairy cattle focused on studies in non-molecular epidemiology. Given the large number of eligible results for dairy (739 articles), we used random sampling to select full-text articles which were reviewed for data availability. The sampling was stratified by publication year (2017 to 2020), with the number of dairy papers selected each year matching the number of available salmonid papers for that year. A pseudo-random number generation function in Microsoft Excel was used for this purpose.

The full text of all articles selected for inclusion was then reviewed and assessed, along with any supplementary

materials and information available on the publisher's Web page for the article. Relevant information for our study (DOI, publication year, species, country, title, type of work and information regarding nature and availability of datasets) was extracted and tabulated. The nature of the data used in the study was classified as commercial (i.e., data collected and managed by commercial producers for their own purposes) or non-commercial data (i.e., data provided by public agencies and industry bodies, data collected from the literature and data collected on purpose for the study such as experimental data or farm survey data). Once completed, the tabulated dataset was verified against each article to identify and correct any extraction error. For salmonids, given the very different nature of datasets between disciplines, molecular epidemiology papers were assessed separately from non-molecular epidemiology papers.

Raw data were considered 'available' when they were provided in the manuscript's main text, as supplementary materials or in an online repository. Data available upon request to the authors were not considered as available in this study. Data considered as raw data in this assessment were non-aggregated data, provided at the level to which they were collected (i.e., excluding any summary data). For mathematical modelling studies (simulation studies), the model inputs and parameter values were considered as the raw data for the purpose of this assessment, not the model outputs such as simulation results for example. The rationale for identifying 'available' datasets in this first stage was that these data are potentially findable and accessible outside of the research team which produced them. In some cases, all the data required to reproduce the results may have been provided in the manuscript itself (either in the text or as tables in the article). Such data could not be evaluated for FAIRness as they did not form a distinct digital resource from the article itself. Therefore, they were excluded from the next stage of the assessment described in the next paragraph.

The information required to assess FAIRness was then extracted for each digital resource identified as available. Furthermore, information about the online repositories used to deposit datasets by the articles' authors was obtained from the Registry of Research Data Repositories ([www.re3data.org](http://www.re3data.org)). A list of 13 FAIR maturity indicators relevant for our work was compiled based on the indicators proposed by Wilkinson et al. [6, 12] (Table 4). Maturity indicators should be applied to a single digital resource, and therefore the target resource must be defined clearly, as some of the principles apply to both the data and the associated metadata [1]. Here, the resources we evaluated were the datasets associated with each paper and identified during the first steps of the work. For molecular epidemiology papers, the datasets of

<sup>1</sup>papers looking at epidemiology-related aspects of sea lice were retained due to high relevance of this species for the aquaculture sector.

interest were those containing genetic data rather than the sample or isolate data. While the authors of this framework have used a binary scoring system (pass or fail) in their maturity indicator tests ([w3id.org/AmIFAIR](http://w3id.org/AmIFAIR)), we felt that some resources presented some maturity regarding a certain indicator while not entirely meeting the requirement. Therefore, resources were assessed with a 3-level scoring system (2, 1 or 0), according to whether they fully met, partially met or did not meet the requirements for each indicator, respectively. After completion of the assessment, the scores were reviewed individually against each digital resource to identify and correct any error.

The assessment of articles and datasets described above was conducted by the first author (AM), while a parallel assessment of 10% of the items included in the study was conducted by one of the co-authors (CF). At the article level, the availability of raw data, data statement and type of data were assessed by CF for 15 randomly selected articles. A minor discordance was noted for 3 of these articles and assigned to the way these items were assessed. The present Methods section was clarified accordingly. At the dataset level, the FAIRness scores were assessed by CF for 8 randomly selected datasets. Different scores were attributed to 8 indicators out of the 104 indicators which were doubly assessed (13 indicators by dataset). Four of these differences were for the indicator I1, and one each for indicators F1.1, F3, A2 and R1. Based on these differences, the maturity levels for these indicators were re-evaluated and re-defined (Table 4) by the two researchers, and all datasets were re-assessed based on these changes.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12917-021-02971-1>.

**Additional file 1.** Literature search flow diagrams.

**Additional file 2.** Full reference list of the literature review.

## Acknowledgements

Not applicable.

## Authors' contributions

AM, CF and MH jointly designed the study. AM conducted the literature review and wrote the manuscript. CF contributed to the literature review. AS and AC supervised the work. All authors contributed to the discussions of the results and the manuscript preparation. The author (s) read and approved the final manuscript.

## Funding

No specific funding was received for this study.

## Availability of data and materials

The dataset supporting the conclusions of this article is available in the FigShare repository under the following collection: <https://doi.org/10.6084/m9.figshare.c.5316863>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 25 February 2021 Accepted: 6 July 2021

Published online: 11 August 2021

## References

1. Wilkinson MD, Dumontier M, IJ J A, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3(1):160018.
2. Jacobsen A, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the European Open Science cloud. *Inf Serv Use*. 2017;37(1):49–56. <https://doi.org/10.3233/ISU-170824>.
3. Jacobsen A, de Miranda AR, Juty N, Batista D, Coles S, Cornet R, et al. FAIR principles: interpretations and implementation considerations. *Data Intell*. 2019;2(1–2):10–29.
4. Thompson M, Burger K, Kaliyaperumal R, Roos M, da Silva Santos LOB. Making FAIR easy with FAIR tools: from creolization to convergence. *Data Intell*. 2019;2(1–2):87–95.
5. van Reizen M, Stokmans M, Basajja M, Ong'ayo A, Kirkpatrick C, Mons B. Towards the tipping point of FAIR implementation. *Data Intell*. 2020;2(1–2):264–75. [https://doi.org/10.1162/dint\\_a\\_00049](https://doi.org/10.1162/dint_a_00049).
6. Wilkinson MD, Sansone S-A, Schultes E, Doorn P, Bonino da Silva Santos LO, Dumontier M. A design framework and exemplar metrics for FAIRness. *Sci Data*. 2018;5(1):180118.
7. van Reizen M, Stokmans M, Mawere M, Basajja M, Ong'ayo AO, Nakazibwe P, et al. FAIR practices in Africa. *Data Intell*. 2019;2(1–2):246–56.
8. Wittenburg P, Lautenschlager M, Thiemann H, Baldauf C, Trilsbeek P. FAIR practices in Europe. *Data Intell*. 2019;2(1–2):257–63.
9. de Miranda AR, Dumontier M. Considerations for the conduction and interpretation of FAIRness evaluations. *Data Intell*. 2019;2(1–2):285–92.
10. FAIRassist. Help you discover resources to measure and improve FAIRness. 2019. Available from: <https://fairassist.org/>. Cited 2020 Oct 18.
11. Research Data Alliance. Results of an Analysis of Existing FAIR Assessment Tools. 2019. Available from: <https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/results-analysis-existing-fair-assessment-tools>. Cited 2020 Oct 18.
12. Wilkinson MD, Dumontier M, Sansone S-A, Bonino da Silva Santos LO, Prieto M, Batista D, et al. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sci Data*. 2019;6(1):174.
13. FAO. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. Rome: Food and Agriculture Organization of the United Nations; 2018.
14. Lafferty KD, Harvell CD, Conrad JM, Friedman CS, Kent ML, Kuris AM, et al. Infectious diseases affect marine fisheries and aquaculture economics. *Annu Rev Mar Sci*. 2015;7(1):471–96. <https://doi.org/10.1146/annurev-marine-010814-015646>.
15. Sapkota A, Sapkota AR, Kucharski M, Burke J, McKenzie S, Walker P, et al. Aquaculture practices and potential human health risks: current knowledge and future priorities. *Environ Int*. 2008;34(8):1215–26. <https://doi.org/10.1016/j.envint.2008.04.009>.
16. Stentiford GD, Sritunyalucksana K, Flegel TW, Williams BAP, Withyachumnarnkul B, Itsathitphaisarn O, et al. New paradigms to help solve the global aquaculture disease crisis. *PLoS Pathog*. 2017 Feb 2;13(2):e1006160. <https://doi.org/10.1371/journal.ppat.1006160>.
17. Brun E. Epidemiology - an important tool for a sustainable aquaculture future. *Prev Vet Med*. 2019;167:158. <https://doi.org/10.1016/j.pvetmed.2019.03.015>.
18. Subasinghe RP. Epidemiological approach to aquatic animal health management: opportunities and challenges for developing countries to increase aquatic production through aquaculture. *Prev Vet Med*. 2005;67(2):117–24. <https://doi.org/10.1016/j.pvetmed.2004.11.004>.

19. Speedy AW. Global production and consumption of animal source foods. *J Nutr.* 2003 Nov 1;133(11):4048S–53S. <https://doi.org/10.1093/jn/133.11.4048S>.
20. Gerber P, Steinfeld H, Henderson B, Mottet A, Opio C, Dijkman J, et al. Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities. Food and Agriculture Organization of the United Nations: Rome, Italy; 2013. 115 p.
21. Broderick GA. Review: Optimizing ruminant conversion of feed protein to human food protein. *Animal.* 2018;12(8):1722–34.
22. Bayliss SC, Verner-Jeffreys DW, Ryder D, Suarez R, Ramirez R, Romero J, et al. Genomic epidemiology of the commercially important pathogen *Renibacterium salmoninarum* within the Chilean salmon industry. *Microb Genomics.* 2018;4(9):e000201.
23. Brosnahan CL, Munday JS, Ha HJ, Preece M, Jones JB. New Zealand rickettsia-like organism (NZ-RLO) and *Tenacibaculum maritimum*: distribution and phylogeny in farmed Chinook salmon (*Oncorhynchus tshawytscha*). *J Fish Dis.* 2019;42(1):85–95. <https://doi.org/10.1111/jfd.12909>.
24. Isla A, Saldarriaga-Córdoba M, Fuentes DE, Albornoz R, Haussmann D, Mancilla-Schulz J, et al. Multilocus sequence typing detects new *Piscirickettsia salmonis* hybrid genogroup in Chilean fish farms: evidence for genetic diversity and population structure. *J Fish Dis.* 2019;42(5):721–37. <https://doi.org/10.1111/jfd.12976>.
25. Söderlund R, Hakhverdyan M, Aspan A, Jansson E. Genome analysis provides insights into the epidemiology of infection with flavobacterium psychrophilum among farmed salmonid fish in Sweden. *Microb Genomics.* 2018;4(12):e000241.
26. Aslam ML, Robledo D, Krasnov A, Moghadam HK, Hillestad B, Houston RD, et al. Quantitative trait loci and genes associated with salmonid alphavirus load in Atlantic salmon: implications for pancreas disease resistance and tolerance. *Sci Rep.* 2020;10(1):1–15.
27. Småge SB, Brevik ØJ, Frisch K, Watanabe K, Duesund H, Nylund A. Concurrent jellyfish blooms and tenacibaculosis outbreaks in northern Norwegian Atlantic salmon (*Salmo salar*) farms. *PLoS One.* 2017;12(11):e0187476. <https://doi.org/10.1371/journal.pone.0187476>.
28. Du X, Bayliss SC, Feil EJ, Liu Y, Wang C, Zhang G, et al. Real time monitoring of *Aeromonas salmonicida* evolution in response to successive antibiotic therapies in a commercial fish farm. *Environ Microbiol.* 2019;21(3):1113–23. <https://doi.org/10.1111/1462-2920.14531>.
29. Büyükekiz AG, Altun S, Hansen EF, Saticioglu IB, Duman M, Markussen T, et al. Infectious pancreatic necrosis virus (IPNV) serotype Sp is prevalent in Turkish rainbow trout farms. *J Fish Dis.* 2018;41(1):95–104. <https://doi.org/10.1111/jfd.12675>.
30. Duman M, Altun S, Cengiz M, Saticioglu IB, Buyuekiz AG, Sahinturk P. Genotyping and antimicrobial resistance genes of *Yersinia ruckeri* isolates from rainbow trout farms. *Dis Aquat Org.* 2017 Jun 19;125(1):31–44. <https://doi.org/10.3354/dao03132>.
31. Johnson-Mackinnon JC, Crosbie PBB, Karlsbakk E, Marcos-Lopez M, Paley R, Nowak BF, et al. Multilocus sequence typing (MLST) and random polymorphic DNA (RAPD) comparisons of geographic isolates of *Neoparamoeba perurans*, the causative agent of amoebic gill disease. *Pathog Basel Switz.* 2019;8(4):244.
32. Karlsen C, Ottem KF, Brevik ØJ, Davey M, Sørum H, Winther-Larsen HC. The environmental and host-associated bacterial microbiota of Arctic seawater-farmed Atlantic salmon with ulcerative disorders. *J Fish Dis.* 2017;40(11):1645–63. <https://doi.org/10.1111/jfd.12632>.
33. Kristoffersen AB, Devold M, Aspehaug V, Gjelstenli O, Breck O, Bang JB. Molecular tracing confirms that infection with infectious pancreatic necrosis virus follows the smolt from hatchery to grow-out farm. *J Fish Dis.* 2018; 41(10):1601–7. <https://doi.org/10.1111/jfd.12844>.
34. la Bastide PY de, Naumann C, Hintz WE. Assessment of intra-specific variability in *Saprolegnia parasitica* populations of aquaculture facilities in British Columbia. *Canada Dis Aquat Organ.* 2018;128(3):235–48. <https://doi.org/10.3354/dao03224>.
35. Sørensen J, Vendramin N, Priess C, Cannimuthu D, Henriksen NH, Iburg TM, et al. Emergence and spread of piscine orthoreovirus genotype 3. *Pathog Basel Switz.* 2020;9(10):823.
36. Adamek M, Hellmann J, Flamm A, Teitge F, Vendramin N, Fey D, et al. Detection of piscine orthoreoviruses (PRV-1 and PRV-3) in Atlantic salmon and rainbow trout farmed in Germany. *Transbound Emerg Dis.* 2019;66(1):14–21. <https://doi.org/10.1111/tbed.13018>.
37. Duman M, Buyuekiz AG, Saticioglu IB, Cengiz M, Sahinturk P, Altun S. Epidemiology, genotypic diversity, and antimicrobial resistance of *Lactococcus garvieae* in farmed rainbow trout (*Oncorhynchus mykiss*). *Iran J Fish Sci.* 2020;19(1):1–18.
38. Jalali SAH, Mohammadinezhad R, Mohammadi A, Latifian MH, Talebi M, Soleimanian-Zad S, et al. Molecular evolution and selection pressure analysis of infectious hematopoietic necrosis virus (IHNV) revealed the origin and phylogenetic relationship of Iranian isolates in recent epidemics in Iran. *Virology.* 2019;535:45–58. <https://doi.org/10.1016/j.virol.2019.06.012>.
39. Jia P, Breyta RB, Li Q, Qian X, Wu B, Zheng W, et al. Insight into infectious hematopoietic necrosis virus (IHNV) in Chinese rainbow trout aquaculture from virus isolated from 7 provinces in 2010–2014. *Aquaculture.* 2018;496:239–46. <https://doi.org/10.1016/j.aquaculture.2018.06.062>.
40. Kibenge MJT, Wang Y, Gayeski N, Morton A, Beardslee K, McMillan B, et al. Piscine orthoreovirus sequences in escaped farmed Atlantic salmon in Washington and British Columbia. *Virol J.* 2019;16(1):41. <https://doi.org/10.1186/s12985-019-1148-2>.
41. Loch TP, Faisal M. Flavobacteria colonizing the early life stages of hatchery-incubated Chinook salmon *Oncorhynchus tshawytscha* (Walbaum 1792) are markedly diverse. *J Fish Dis.* 2018;41(5):829–45. <https://doi.org/10.1111/jfd.12795>.
42. Terceti MS, Vences A, Matanza XM, Dalsgaard I, Pedersen K, Osorio CR. Molecular epidemiology of *Photobacterium damsela* subsp *damsela* outbreaks in marine rainbow trout farms reveals extensive horizontal gene transfer and high genetic diversity. *Front Microbiol.* 2018;9:2155. <https://doi.org/10.3389/fmicb.2018.02155>.
43. Vennerström P, Välimäki E, Lyytikäinen T, Hautaniemi M, Vidgren G, Koski P, et al. Viral haemorrhagic septicaemia virus (VHSV id) infections are detected more consistently using syndromic vs. active surveillance. *Dis Aquat Org.* 2017;126(2):111–23. <https://doi.org/10.3354/dao03161>.
44. Gallagher MD, Karlsen M, Petterson E, Haugland O, Matejusova I, Macqueen DJ. Genome sequencing of SAV3 reveals repeated seeding events of viral strains in norwegian aquaculture. *Front Microbiol.* 2020;11:740.
45. Rodríguez FH, Flores-Mara R, Yoshida GM, Barria A, Jedlicki AM, Lhorente JP, et al. Genome-Wide Association Analysis for Resistance to Infectious Pancreatic Necrosis Virus Identifies Candidate Genes Involved in Viral Replication and Immune Response in Rainbow Trout (*Oncorhynchus mykiss*). *G3 Bethesda Md.* 2019;9(9):2897–904.
46. Barker SE, Bricknell IR, Covelto J, Purcell S, Fast MD, Wolters W, et al. Sea lice, *Lepeophtheirus salmonis* (Krøyer 1837), infected Atlantic salmon (*Salmo salar* L.) are more susceptible to infectious salmon anemia virus. *Plos One.* 2019;14(1):e0209178.
47. Daros RR, Eriksson HK, Weary DM, von Keyserlingk MAG. The relationship between transition period diseases and lameness, feeding time, and body condition during the dry period. *J Dairy Sci.* 2020;103(1):649–65. <https://doi.org/10.3168/jds.2019-16975>.
48. Escobar LE, Escobar-Dodero J, Phelps NBD. Infectious disease in fish: global risk of viral hemorrhagic septicemia virus. *Rev Fish Biol Fish.* 2018;28(3):637–55. <https://doi.org/10.1007/s11160-018-9524-3>.
49. Guerrero-Cabrera L, Luna Olivera BC, Villavicencio-Pulido JG, Ortiz Luna RJ. Proximity and density of neighboring farms and water supply, as risk factors for bacteriosis: a case study of spatial risk analysis in tilapia and rainbow trout farms of Oaxaca. *Mexico Aquaculture.* 2020;520:734955. <https://doi.org/10.1016/j.aquaculture.2020.734955>.
50. Myksovoll MS, Sandvik AD, Albretsen J, Asplin L, Johnsen IA, Karlsen Ø, et al. Evaluation of a national operational salmon lice monitoring system—from physics to fish. *PLoS One.* 2018;13(7):e0201338. <https://doi.org/10.1371/journal.pone.0201338>.
51. Nekouei O, Vanderstichel R, Kaukinen KH, Thakur K, Ming T, Patterson DA, et al. Comparison of infectious agents detected from hatchery and wild juvenile Coho salmon in British Columbia, 2008–2018. *PLoS One.* 2019;14(9):e0221956. <https://doi.org/10.1371/journal.pone.0221956>.
52. Riihã V, Sundberg L-R, Ashrafi R, Hyvärinen P, Karvonen A. Rearing background and exposure environment together explain higher survival of aquaculture fish during a bacterial outbreak. *J Appl Ecol.* 2019;56(7):1741–50. <https://doi.org/10.1111/1365-2664.13393>.
53. Samsing F, Johnsen I, Trembl EA, Dempster T. Identifying “firebreaks” to fragment dispersal networks of a marine parasite. *Int J Parasitol.* 2019;49(3–4):277–86. <https://doi.org/10.1016/j.ijpara.2018.11.005>.
54. Soler-Jiménez LC, Paredes-Trujillo AI, Vidal-Martínez VM. Helminth parasites of finfish commercial aquaculture in Latin America. *J Helminthol.* 2017;91(2):110–36. <https://doi.org/10.1017/S0022149X16000833>.
55. Palotay JL. Snomed-Snovet: an information system for comparative medicine. *Med Inform (Lond).* 1983;8(1):17–21. <https://doi.org/10.3109/14639238309010917>.

56. Rodríguez-Iglesias A, Rodríguez-González A, Irvine AG, Sesma A, Urban M, Hammond-Kosack KE, et al. Publishing FAIR Data: An Exemplar Methodology Utilizing PHI-Base. *Front Plant Sci.* 2016;7:641. Available from: <https://www.frontiersin.org/articles/10.3389/fpls.2016.00641/full>. Cited 2020 Nov 22
57. Houtkoop BL, Chambers C, Macleod M, Bishop DVM, Nichols TE, Wagenmakers E-J. Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Adv Methods Pract Psychol Sci.* 2018;1(1):70–85. <https://doi.org/10.1177/2515245917751886>.
58. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One.* 2015 Aug 26;10(8):e0134826. <https://doi.org/10.1371/journal.pone.0134826>.
59. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. *PLoS One.* 2011;6(6):e21101. <https://doi.org/10.1371/journal.pone.0021101>.
60. Tenopir C, Christian L, Allard S, Borycz J. Research data sharing: practices and attitudes of geophysicists. *Earth Space Sci.* 2018;5(12):891–902. <https://doi.org/10.1029/2018EA000461>.
61. Tenopir C, Rice NM, Allard S, Baird L, Borycz J, Christian L, et al. Data sharing, management, use, and reuse: practices and perceptions of scientists worldwide. *PLoS One.* 2020;15(3):e0229003. <https://doi.org/10.1371/journal.pone.0229003>.
62. Dai S-Q, Li H, Xiong J, Ma J, Guo H-Q, Xiao X, et al. Assessing the extent and impact of online data sharing in Eddy covariance flux research. *J Geophys Res Biogeosci.* 2018;123(1):129–37. <https://doi.org/10.1002/2017JG004277>.
63. Hipsley CA, Sherratt E. Psychology, not technology, is our biggest challenge to open digital morphology data. *Sci Data.* 2019;6(1):41. <https://doi.org/10.1038/s41597-019-0047-0>.
64. Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLoS Biol.* 2018;16(11):e2006930. <https://doi.org/10.1371/journal.pbio.2006930>.
65. Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in The BMJ and PLOS Medicine. *BMJ.* 2018;360:k400. <https://doi.org/10.1136/bmj.k400>.
66. Byrd JB. Responsible, practical genomic data sharing that accelerates research. *Nat Rev Genet.* 2020;21(10):615–29. <https://doi.org/10.1038/s41576-020-0257-5>.
67. Dallmeier-Tiessen S, Darby R, Gitmans K, Lambert S, Matthews B, Mele S, et al. Enabling sharing and reuse of scientific data. *New Rev Inf Netw.* 2014;19(1):16–43. <https://doi.org/10.1080/13614576.2014.883936>.
68. Perrier L, Blondal E, MacDonald H. The views, perspectives, and experiences of academic researchers with data sharing and reuse: a meta-synthesis. *PLoS One.* 2020;15(2):e0229182. <https://doi.org/10.1371/journal.pone.0229182>.
69. Kim Y, Burns CS. Norms of data sharing in biological sciences: The roles of metadata, data repository, and journal and funding requirements. *J Inf Sci.* 2016;42(2):230–45. <https://doi.org/10.1177/0165551515592098>.
70. Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: the Belmont Forum’s open data survey. *PLoS One.* 2016;11(1):e0146695. <https://doi.org/10.1371/journal.pone.0146695>.
71. Vines TH, Andrew RL, Bock DG, Franklin MT, Gilbert KJ, Kane NC, et al. Mandated data archiving greatly improves access to research data. *FASEB J.* 2013;27(4):1304–8. <https://doi.org/10.1096/fj.12-218164>.
72. Bloemers M, Montesanti A. The FAIR funding model: providing a framework for research funders to drive the transition toward FAIR data management and stewardship practices. *Data Intell.* 2019;2(1–2):171–80.
73. EC. Final Report and Action Plan from the European Commission Expert Group on FAIR Data: Turning Fair into Reality. Brussels: European Commission; 2018. 78 p. Available from: <https://doi.org/0.2777/1524>
74. Gautam R, Vanderstichel R, Boerlage AS, Revie CW, Hammell KL. Effect of timing of count events on estimates of sea lice abundance and interpretation of effectiveness following bath treatments. *J Fish Dis.* 2017;40(3):367–75. <https://doi.org/10.1111/jfd.12519>.
75. Meyer A, Bannister-Tyrell M, Mackenzie C, Stegeman A, Cameron A. Barriers to the adoption of a fish health data integration initiative in the Chilean salmonid production. *Comput Electron Agric.* 2020;179:105853. <https://doi.org/10.1016/j.compag.2020.105853>.
76. Happold J, Meyer A, Sadler R, Cowled B, Mackenzie C, Gallardo Lagno A, et al. Effectiveness of antimicrobial treatment of salmonid rickettsial septicaemia in commercial salmon and trout farms in Chile. *Aquaculture.* 2020;525:735323. <https://doi.org/10.1016/j.aquaculture.2020.735323>.
77. Happold J, Sadler R, Meyer A, Cowled B, Mackenzie C, Gallardo Lagno A, et al. Effectiveness of vaccination for the control of salmonid rickettsial septicaemia in commercial salmon and trout farms in Chile. *Aquaculture.* 2020;520:734968. <https://doi.org/10.1016/j.aquaculture.2020.734968>.
78. Haine D, Delgado H, Cue R, Sewalem A, Wade K, Lacroix R, et al. Culling from the herd’s perspective—exploring herd-level management factors and culling rates in Québec dairy herds. *Prev Vet Med.* 2017;147:132–41. <https://doi.org/10.1016/j.prevetmed.2017.08.020>.
79. Holub P, Kohlmayer F, Prasser F, Mayrhofer MT, Schlünder I, Martin GM, et al. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-health. *Biopreservation Biobanking.* 2018;16(2):97–105. <https://doi.org/10.1089/bio.2017.0110>.
80. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol.* 2009;62(10):1006–12. <https://doi.org/10.1016/j.jclinepi.2009.06.005>.

## Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

