

# Unleashing the power of administrative health data: the Scottish model

Stephen Pavis<sup>a,b</sup> and Andrew D Morris<sup>a</sup>

<sup>a</sup> Farr Institute of Health Informatics Research, Edinburgh, UK

<sup>b</sup> Corresponding author: [s.pavis@nhs.net](mailto:s.pavis@nhs.net)

## Article history

Publication date: September 2015

Citation: Pavis S, Morris AD. Unleashing the power of administrative health data: the Scottish model. *Public Health Res Pract*. 2015;25(4):e2541541. doi: <http://dx.doi.org/10.17061/phrp2541541>

## Key points

- Government administrative information is a potentially valuable untapped resource for health service improvement; however, the public's trust must be established before it is used for research
- The 'Scottish model' ensures privacy protection via 'safe havens', while at the same time providing data access to approved researchers, who are held accountable for privacy protection. Commercial organisations can access data through collaboration with public entities
- During the past 3 years, more than 150 projects have been completed by industry, academia and health service provider researchers using this model

## Abstract

Data and information generated through the provision and administration of health and social care provide potentially valuable untapped resources that can contribute to the development of effective and efficient services. We describe the Scottish system, which seeks to unleash, at scale, the power of administrative and health service data as part of the UK-wide Farr Institute of Health Informatics Research program. The 'Scottish model' balances current public attitudes and views around the use of administrative and health data for research purposes with researchers' data requirements, and does so within Scotland's legal framework. The past 3 years has seen the completion of more than 150 projects by researchers from industry (17%), academia (53%) and health service providers (30%). In the future, the aim will be to ensure that research findings are disseminated widely and used to both improve health service provision and further develop public trust.

## Introduction

Health and social care services internationally need to be optimised to meet future demand. Arguably, this requires a closer coupling of service delivery and research, whereby innovations in services can be rapidly tested, evaluated and fed back into a continual process of service optimisation and improvement. Data and information generated through the provision and administration of health and social care provide potentially valuable untapped resources that can contribute to the development of effective and efficient services. In this paper, we describe the Scottish system, which seeks to unleash, at scale, the power of health service and wider administrative data as part of a UK-wide Farr Institute of Health Informatics Research program. The term 'administrative data' is used here to refer to information that is collected about citizens during the provision of public services – for example, by state-provided education and housing – and the taxation system.

## Scotland's healthcare system and demography

Scotland has a population of 5 million. Healthcare is delivered by a single state-provided National Health Service (NHS), administered via the Scottish Government. Both primary and secondary care are funded via general taxation and are free to patients at the point of use, including all medicines. The only areas where nominal charges are made to patients are for dental treatment and eyewear (but not eye examinations). NHS providers of secondary care (primarily hospitals) receive annual funding from Regional Health Boards. Resources for the provision of primary care are allocated to general practitioners or family doctors who 'contract' to work for the NHS and receive payments based on a weighted capitation formula, which takes account of variation in patient lists (e.g. demographics, levels of deprivation, rurality), plus additional payments for certain specified activities (e.g. care processes and outcomes for certain diseases, such as diabetes and coronary heart disease).

The population is relatively stable with comparatively low levels of geographic mobility, and there is very little private healthcare provided (<2%). Health outcomes in Scotland are relatively poor when compared with developed nations, with mortality in working-age populations comparatively high, and mortality for certain key diseases (e.g. circulatory disease and several cancers) higher than in most other European countries.<sup>1</sup>

## Scotland's administrative data – a national treasure

Scotland has some of the best administrative and care data in the world. Within the NHS, data have been collected at the national level for more than 40 years, with the Information Services Division of National Services

Scotland charged with ensuring completion, quality and comparability across geographic regions and health services. Regional Health Boards (14 in Scotland) hold further clinical data that can be used to create rich phenotypes.

The Community Health Index (CHI) is a register of all patients who use the Scottish NHS. The register ensures that patients can be correctly identified, and that all information pertaining to a patient's health is available to care providers. Patients are identified using a 10-digit number – the CHI number. It is estimated that between 96.5% and 99.9% of the Scottish population have a CHI number. This unique patient identifier allows healthcare records for individuals to be linked across time and location, and is critical to achieving efficient data linkage and patient databases analyses.

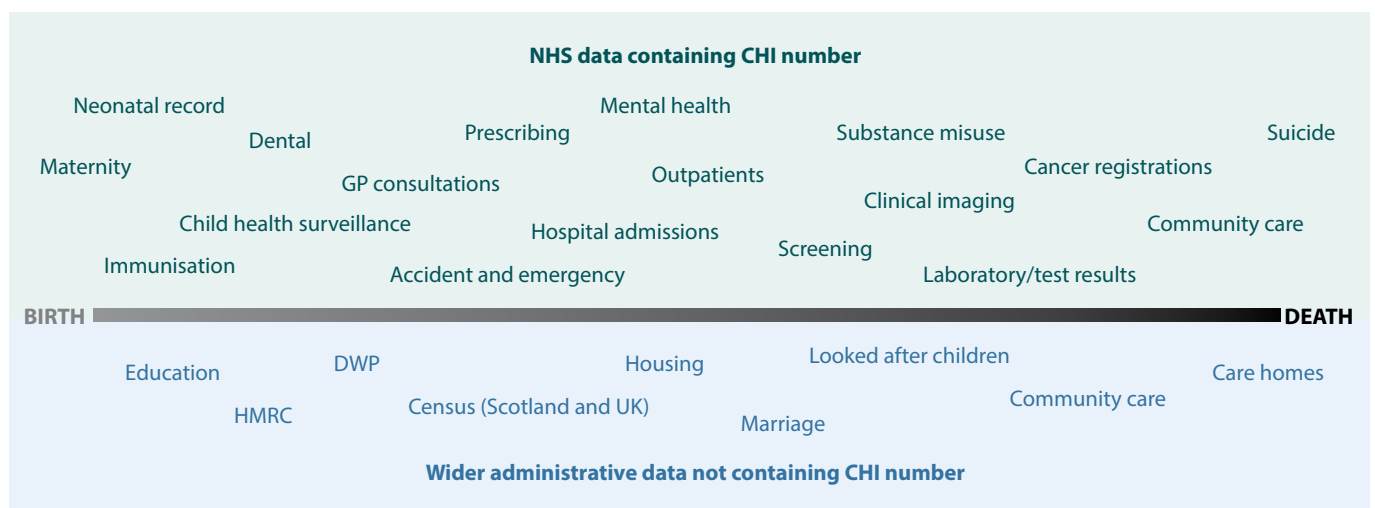
Wider administrative data from the economic and social spheres are available to supplement health data and provide potential resources to understand the social influences on health, illness, service use and outcomes, as well as (in)equality and social mobility across regions and historical periods. Although these records do not contain the CHI number, a process has been developed that allows linkage of health and nonhealth information to create de-identified research datasets.

Types of administrative data and other data that use the CHI number are shown in Figure 1.

## Public acceptance and the legal basis for using health data for research

Excellent and scalable informatics research requires more than simply the existence of excellent data and strong researchers. It also requires sound administrative processes and supporting infrastructures. In turn, these must operate within the country's legal framework and

**Figure 1. National-level data resources**



CHI = Community Health Index; DWP = Department for Works and Pensions; HMRC = HM Revenue and Customs; NHS = National Health Service

have the support of the public, particularly since gaining each individual's consent to use administrative data (particularly historical data) is close to impossible. Data must be processed efficiently, and IT provision must be powerful and responsive to researchers' requirements. In these respects, Scotland benefited from the Scottish Health Informatics Programme (funded by the Wellcome Trust) and subsequently the Farr Institute of Health Informatics Research (funded by the Medical Research Council and nine other funders). These facilitated public engagement programs about the acceptability of using nonconsented public data and the development of ways to link data for research while also protecting individuals' privacy. Resources were also given to purchase a secure high-performance computing environment.

## Building trusted and transparent systems

A variety of social science methods have been used to understand the public's views about the use of administrative health and care data, including focus groups, and public workshops and panels in which researchers talked about how electronic data can support different types of research. In addition, as early as 2002, the Scottish Government's Chief Scientist Office created a Public Involvement Group to ensure public representation in the research arena. The messages from public engagement are reasonably consistent and show that citizens tend to offer 'conditional support' for the use of administrative and health data in research. The public understands that gaining consent from each individual for every study is impractical and, under these circumstances, points to the importance of "transparency, de-identifying data, ensuring that IT systems are secure" and that access is limited to 'trusted personnel', who conduct research that has 'public benefit'. Public engagement suggests that trust is highest in health service workers and university academics, with less trust shown towards commercial companies. Government employees appear to lie somewhere between the NHS and academic employees, and those of commercial organisations.<sup>2</sup>

## Legal considerations

The primary legislation governing use of personal data in Scotland is the Data Protection Act 1998. This Act places various duties on 'data controllers' to process 'personal' and 'sensitive personal data' in what are termed 'fair' and 'lawful' ways. With regard to the use of data for research purposes, section 33 of the Act allows data to be used as long as they are not processed to "support measures or decisions with respect to particular individuals", or "in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject". To ensure compliance with the law and maintain public

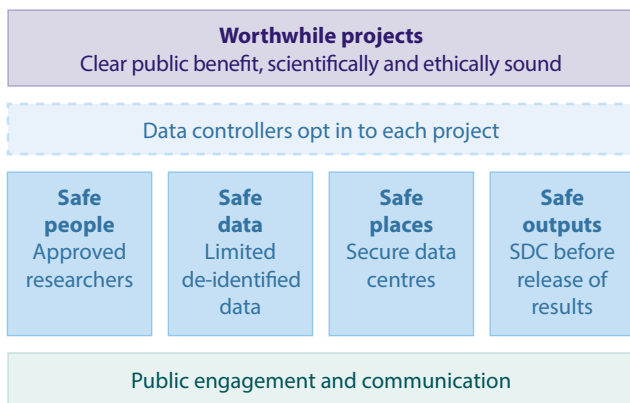
trust and transparency, the NHS has a system of Caldicott Guardians<sup>3</sup> at the local regional level and a Public Benefits and Privacy Panel at the national level.<sup>4</sup> In combination, these administrative structures ensure that data are only made available when it is legal to do so and when there is clear public benefit.

## The 'Scottish model'

The Scottish model has been developed against the background of 'conditional public consent' and the legal framework. The key elements of the Scottish approach are (Figure 2):

1. Scotland does not have nor seeks to achieve a single data warehouse that can meet all research needs. Instead, data, under the responsibility of different data controllers, are held at both regional and national levels, and subsets (groups of variables) are brought together when there are clear research questions that have public benefit. Access to nonconsented data requires consent from the legal data controllers on a project-by-project basis. The data controllers must actively assess research proposals and assure themselves that they are both in the public interest and meet legislative requirements.
2. De-identified research datasets are provided to researchers through a federated network of 'safe havens'. These environments operate to high levels of security and are accredited by NHS Scotland. Researchers access and analyse data through virtual private networks, but cannot remove data (or analyses) before research outputs have been checked to ensure they do not identify individuals or breach privacy.
3. Only 'approved researchers' are provided with access to research datasets. Approved researchers must undergo a short course to ensure adequate understanding of legal frameworks and privacy risks, and work within a public sector organisation. Contracts are put in place to hold both the researcher and their institution accountable for behaviours while accessing data.
4. Support to the private and commercial sector is vital, and much research undertaken by commercial organisations has clear public benefit. However, the public remains apprehensive about commercial organisations directly accessing health data (even when it is de-identified). Trust must therefore be developed over time and through the demonstration of good practice and public benefit. Accordingly, cross-sector partnerships are actively pursued, with commercial partners setting research questions and driving research forward through project steering groups.

**Figure 2.** Securing trust: data controllers and the public



SDC = statistical disclosure control

## Creating research datasets while protecting individuals' privacy

The decision to create bespoke, project-specific research datasets, rather than permanently linked data held in a single warehouse, was driven by legislative requirements and our understanding of public attitudes. The resulting system appears to have public support, but brings with it certain challenges. Data are regularly moved between various data controllers and this creates potential security issues. This risk is mitigated by a combination of modern secure data transfer technologies and what has become known as 'linkage using a separation of functions'. This approach rests on a clear segregation between an organisation that links individuals together across multiple datasets, and an organisation that holds and provides a linked, de-identified research dataset.

When approval has been granted to create a linked research dataset, data controllers first send identifying information (names, dates of birth, addresses) to a 'trusted third party' (TTP), who uses this information to link individuals who appear in more than one dataset. In Scotland, this is facilitated by the CHI register, but nonhealth datasets that do not contain the CHI number can also be linked using probabilistic linkage techniques that match other identifying information (e.g. names, addresses or dates of birth). It is important to note that the TTP does not receive the data to be used in the research study (i.e. personal health information), only the identifiers for study participants. Whether linkage is through direct matching using the CHI or probabilistic techniques using other identifying information, the TTP provides a project-specific 'index number' for each individual and returns this to the data controllers. The index number used is not the same for each data controller or across projects, meaning that data controllers cannot link data together without the system. The data controllers then send the index number and information required for the study (e.g. health or education data in the example in Figure 3)

to a second separate organisation, which then joins the datasets using a 'key' provided by the TTP. De-identified data are then provided to the approved researcher within a secure analytic platform. In this model, the TTP knows the identities of individuals they have indexed but nothing about their health records, while the second organisation knows the research variables (e.g. health or education outcomes) but not individuals' identities. In combination, these two separate organisations comprise the 'safe haven'. These processes are presented in Figure 3.

## Support to researchers – eData Research and Innovation Service (eDRIS)

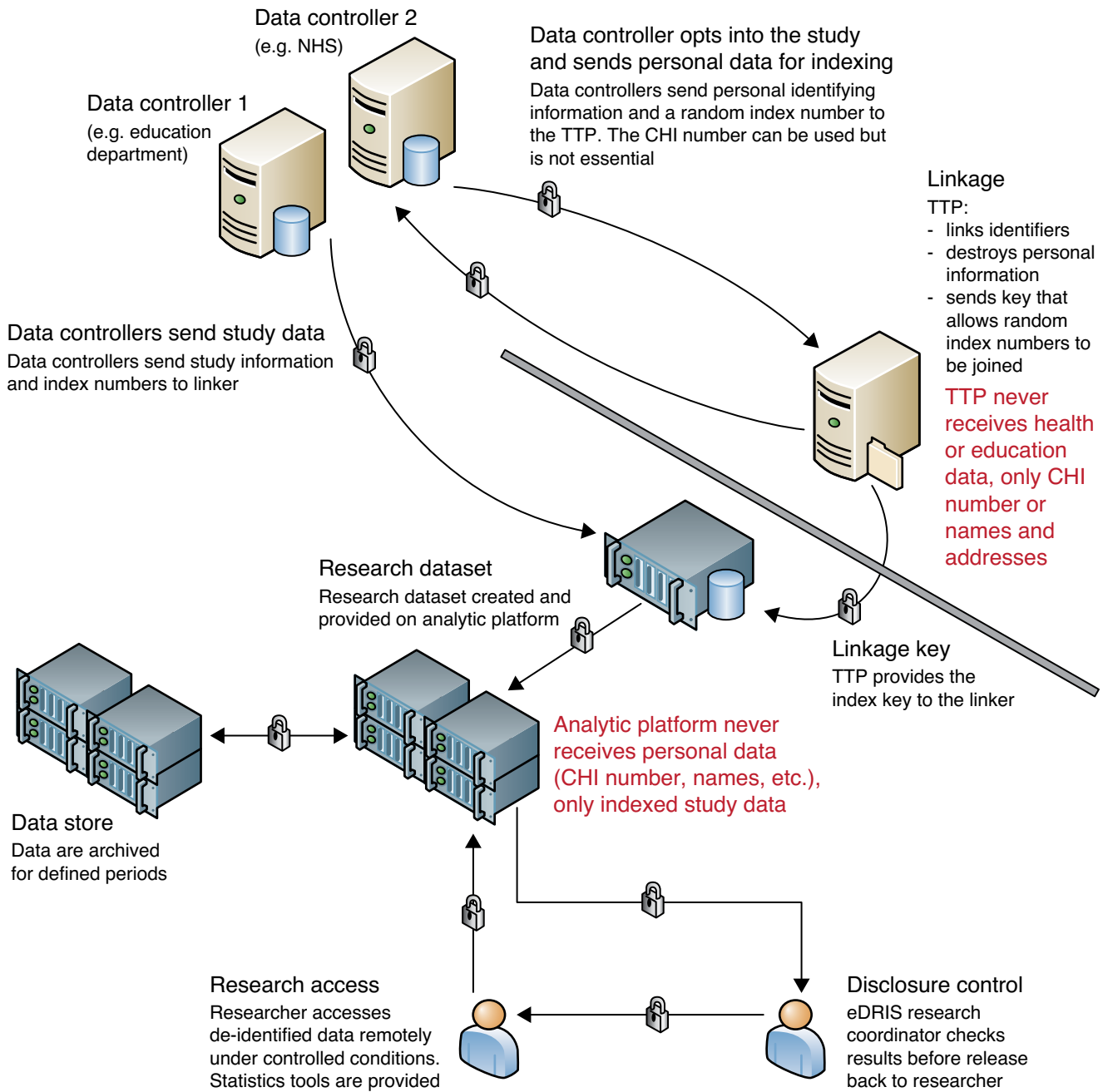
The Scottish approach to providing linked research data for health informatics involves multiple organisations and distributed data, managed by various data controllers, some within and some outside the NHS. The eDRIS researcher support service was introduced in 2013 to help researchers navigate the system and understand what data are available, their quality and completeness, and the processes and procedures required before accessing them.<sup>4</sup> The eDRIS team sits between the researchers, the data controllers, the TTP and the secure analytic environment, and aims to ensure efficient communication and coordination between parties as datasets are created and provided. At the start of a project, the researcher is allocated a named 'research coordinator', who works with the researcher throughout the project. The research coordinator provides advice on potential sources of information; coding and terminology issues; where datasets are located and the process for accessing them; and achieving 'approved researcher' status. They also liaise with data controllers so that data are indexed and provided on a secure analytic platform, and undertake statistical disclosure control before the results are released from the secure environment.

Recently, the UK's Medical Research Council provided funds to create the Farr Institute of Health Informatics Research.<sup>5</sup> In Scotland, a consortium of six universities and the NHS have come together to further develop and improve the Scottish model. An important and innovative development is its co-location, within a single building of the eDRIS support service, with academics (epidemiologists and social scientists) and computer scientists. The aim is to develop further cross-sectoral and cross-disciplinary relationships within a research-friendly environment, and ultimately to improve research outputs.

## Discussion

The Scottish model balances current public attitudes and views around the use of administrative and health data for research purposes with researchers' data

**Figure 3.** Scottish data linkage using the principle of separation of functions



CHI = Community Health Index; eDRIS = eData Research and Innovation Service; NHS = National Health Service; TTP = trusted third party

requirements and Scotland's legal framework. During the past 3 years, more than 150 projects have been completed by researchers from industry (17%), academia (53%) and health service providers (30%). Scottish administrative data have also been marshalled within a range of study designs. For example, Pell et al. used a natural experimental design to understand the effects of the ban on tobacco use in public places on acute coronary syndrome<sup>6</sup> and childhood asthma.<sup>7</sup> Livingston et al.<sup>8</sup> used a retrospective cohort analysis to understand the relationship between type 1 diabetes and the risk of cardiovascular disease, while Moon et al.<sup>9</sup> used Scottish drug data in combination with other European countries to

look at the potential cost savings from prescribing generic rather than labelled drugs.

In the future, the aim will be to ensure that research findings are disseminated widely and used to improve health service provision and further develop public trust. In time, we hope to allow trusted researchers from commercial organisations to have direct access to de-identified research datasets. Our experiences suggest that it is only when there are high levels of trust between the public, researchers and those providing data linkage services that the full potential of health data can be realised.

## Competing interests

None declared

## Author contributions

SP drafted the manuscript and revised it following reviewers' comments. ADM reviewed the manuscript and completed the second draft.

## References

1. Whyte B, Ajetunmobi T. Still "the sick man of Europe"? Scottish mortality in a European context 1950–2010: an analysis of comparative mortality trends. Glasgow: Glasgow Centre for Population Health, ScotPHO; 2012 [cited 2015 Aug 11]. Available from: [www.scotpho.org.uk/publications/reports-and-papers/937-still-the-sick-man-of-europe-scottish-mortality-in-a-european-context-1950-2010-an-analysis-of-comparative-mortality-trends](http://www.scotpho.org.uk/publications/reports-and-papers/937-still-the-sick-man-of-europe-scottish-mortality-in-a-european-context-1950-2010-an-analysis-of-comparative-mortality-trends)
2. Aitken M. SHIP public engagement: summary of focus group findings. Edinburgh: Scottish Health Informatics Programme; 2011 [cited 2015 May]. Available from: [www.scot-ship.ac.uk/sites/default/files/Reports/Focus\\_Group\\_Findings\\_Briefing\\_Paper.pdf](http://www.scot-ship.ac.uk/sites/default/files/Reports/Focus_Group_Findings_Briefing_Paper.pdf)
3. National Health Service Scotland. Caldicott Guardians. Edinburgh: National Health Service Scotland; 2015 [cited 2015 May]. Available from: [www.nhs.uk/nhs.uk/pages/corporate/caldicott\\_guardians.php](http://www.nhs.uk/nhs.uk/pages/corporate/caldicott_guardians.php)
4. NHS Scotland. The public benefit and privacy panel for health and social care. Edinburgh: NHS Scotland; 2015 [cited 2015 May]. Available from: [www.informationgovernance.scot.nhs.uk](http://www.informationgovernance.scot.nhs.uk)
5. Farrinstitute.org [Internet]. Edinburgh: The Farr Institute of Health Informatics Research; c2013–15 [cited 2015 May]. Available from: [www.farrinstitute.org](http://www.farrinstitute.org)
6. Pell J, Haw S, Cobbe S, Newby DE, Pell AC, Fischbacher C, et al. Smoke-free legislation and hospitalization for acute coronary syndrome. *N Engl J Med*. 2008;359(5):482–91.
7. Mackay D, Haw S, Ayres JG, Fischbacher C, Pell JC. Smoke-free legislation and hospitalizations for childhood asthma. *N Engl J Med*. 2010;363(12):1139–45.
8. Livingston SJ, Looker HC, Hothersall EJ, Wild SH, Lindsay RS, Chalmers J, et al. Risk of cardiovascular disease and total mortality in adults with type 1 diabetes: Scottish registry linkage study. *PLoS Med*. 2012;9(10):e1001321.
9. Moon JC, Godman B, Pezold M, Alvarez-Madrado S, Bennett K, Bishop I, et al. Different initiatives across Europe to enhance losartan utilization post generics: impact and implications. *Front Pharmacol*. 2014;5:219.

Copyright: 

© 2015 Pavis and Morris. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence, which allows others to redistribute, adapt and share this work non-commercially provided they attribute the work and any adapted version of it is distributed under the same Creative Commons licence terms. See: [www.creativecommons.org/licenses/by-nc-sa/4.0/](http://www.creativecommons.org/licenses/by-nc-sa/4.0/)