D4.3 Report on the usage and auditing data hubs indicators and proposal of an incentive system Version 2.0

Document Information

Contract Number	965345
Project Website	http://www.healthycloud.eu/
Contractual Deadline	M27, May 2023
Dissemination Level	PU - Public
Nature	R - Report
Author(s)	Teresa D'Altri (CRG) Laura Portell Silva (BSC) Celia Alvarez-Romero (SAS) Inês Lourenço (SPMS) Vanessa Mendes (SPMS) Cátia Pinto (SPMS) Andrea Mero Vega (CRG)
Contributor(s)	Salvador Capella-Gutierrez (BSC) Lidia López (BSC) Eva García (BBMRI) Michaela Mayrhofer (BBMRI) Stefan Mathis-Edenhofer (GÖG)
Reviewer(s)	Mark Dietrich (EGI) Lorenz Dolanski-Aghamanoukjan (GÖG)
Keywords	Data usage, auditing, indicators, data hubs, incentives, data access, data sharing, health research

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Change Log

Version	Author	Date	Description of Change
V0.1	Teresa D'Altri	1/03/2023	Initial Draft- outline
V0.2	Teresa D'Altri and Laura Portell	31/03/2023	Comprehensive draft including contribution from partners gathered during collaborative sessions
V1.0	Teresa D'Altri and Laura Portell	14/04/2023	Final draft shared with WP4 partners
v1.1	Teresa D'Altri and Laura Portell	27/04/2023	Improved draft shared with the reviewers
v2.0	Teresa D'Altri and Laura Portell	29/05/2023	Final document including improvements advised by reviewers
			(Final Change Log entries reserved for releases to the EC)

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Table of contents

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Executive Summary

HealthyCloud WP4 'Experiences on health data management: national, regional and domain-specific data hubs' focuses on how health data is managed by the data hubs.

Data hubs play and will continue to play an important role in guaranteeing that health data is accessed, analyzed, shared and reused following the general principles of a framework that is ethically sound and legally compliant, as outlined by HealthyCloud WP2 'aal, Legal and Societal impact of cross-border health data access for cloud analysis'.

Data hubs are also well-positioned to identify and elaborate on the different incentives that drive individuals and organizations to deposit data in and share it through those hubs.

In this sense, deliverable D4.3 'Report on the usage and auditing data hubs indicators and proposal of an incentive system' presents a comprehensive study of both usage and auditing data hubs indicators, and existing incentive systems. The conclusions of this document include a proposal of an incentive system to promote data sharing in the health research field.

1. Introduction

HealthyCloud WP4 focuses on how health data is managed by the data hubs [1]. Previous results of WP4 were presented in deliverables:

- D4.1 'Recommendations for integration in HealthyCloud, including an analysis of data hub patterns of governance' included the analysis of a dedicated survey performed, through the stratification of the results. As results of D4.1, patterns of data hub governance were represented and recommendations for integration in HealthyCloud were shown in the conclusions.
- D4.2 'Report on current discoverability solutions and FAIR adoption level' analyzed the survey responses related to the FAIR principles, including stratification of the results, to generate a set of best practices for enabling discoverability of data collections within a data hub at different FAIRness levels.

In this context, and since data sharing and reuse is the main reason for existence for data hubs, deliverable D4.3 presents the outcomes of the analysis carried out to propose a set of incentives for the future HealthyCloud ecosystem, aiming to promote the data sharing in health research through those hubs.

By conducting this analysis, outputs of HealthyCloud WP2 are taken into account to analyze the legal and ethical requirements for health data accessing, processing, sharing and reuse.

In connection with HealthyCloud WP3, the incentives to structure the underlying data collections following the FAIR data principles will be analyzed. This specific analysis will focus on the balance of a priori efforts vs. expected benefits as this can be a driver to improve existing models and propose new ones.

Finally, information associated with the incentive models will be incorporated into the meta-catalogue proposed in HealthyCloud WP6. It will contribute to having a broader understanding of how health data can be made available and interoperable.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

The importance of metrics in data hubs

Due to numerous challenges such as ensuring transparency, traceability, immutability and legal compliance, the management of health data has become increasingly complex and challenging [2]. In this regard, it is crucial to maintain the accuracy and reliability of data while ensuring that information can be easily accessed when needed.

Data hubs are designed to centralize data storage and discovery of health datasets from different sources, and for that reason, metrics are essential and play a crucial role in effectively monitoring the performance and impact of these activities [2].

A metric is a tool specifically designed to facilitate effective decision-making, enhance performance, and promote accountability within an organization [3]. It can be translated into a quantitative measurement or indicator used to evaluate and track performance, including efficiency, effectiveness, quality, productivity, and profitability, as well as the progress or success of a particular activity, project, or system. By collecting and analyzing relevant performance-related data, data hubs can gain valuable insights into the effectiveness of their data-sharing initiatives and identify areas for improvement. The use of metrics in data hubs promotes transparency, accountability, and data-driven decision-making, which are critical for achieving the goals of organizations and driving meaningful change in the communities they serve.

Metrics and incentive systems are intrinsically linked when it comes to the success of health data hubs. Metrics provide a way to track and analyze data, which can be used to identify areas for improvement and to benchmark success factors, but Incentive systems, on the other hand, provide a way to improve those measures. By establishing clear metrics for success and tying incentives to the achievement of those metrics, health data hubs can encourage collaboration and participation from a variety of stakeholders.

Exploring the vital role of incentive systems in driving health data hubs

The success of data hubs is largely dependent on the quality and quantity of the data collected. Incentive systems can play a vital role in driving the success of health data hubs, as they not only encourage data sharing but also help to build trust between individuals and the organizations managing the data for research and innovation.

Incentives are designed to inspire or motivate individuals to perform a particular action, ultimately influencing their behavior or performance. These stimuli usually take the form of tangible rewards such as bonuses, promotions, or gifts, and are generally separate from the core functionality of data hubs [3].

In order to promote data sharing in the health research field, it is essential to establish an effective incentive system that encourages researchers to share their data. To fully understand the impact of these incentive systems, a thorough analysis was conducted to identify a set of incentives that can encourage researchers and institutions to participate in data sharing through data hubs. This analysis will be further enhanced by the methodology proposed in the next section, which will provide a comprehensive approach to understanding the impact of metrics and incentive systems in promoting data sharing.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

2. Methodology

2.1. Conceptual set up of the survey

The common methodology selected to gather information for both data hubs usage metrics and incentive systems was a survey targeting the existing data hubs. This was defined in the description of the action, as the most effective way to collect data from such a young and fast-changing ecosystem.

As presented in the introduction, WP4 has already carried out a survey to complete tasks 4.1 "Analysis of existing data hubs governance models" and 4.2 "Analysis of the data hubs operation related to the reference guidelines defined in WP3". In order to improve the process, the answers to the former survey were used to narrow the list of recipient data hubs to focus only on those that are relevant to the topic of this deliverable, in order to analyze a collection of all significant answers. This approach reduced the number of answers, yet increased their analytical value. The list of 49 targeted data hubs is provided as annex 1, where the 24 who answered are marked.

The analysis of the results was carried out in a series of collaborative working sessions with the WP partners.

2.2. Auditing of data hubs' usage

In order to evaluate how the European Data hubs currently measure their usage, we designed a set of five questions (plus an open comment field) listed in the second part of annex 2. A partner of WP4, the European Genome-phenome Archive, is a data hub itself, thus could contribute its experience to the design of the questions.

The questions deal with the following aspects:

- metrics for data usage (measurements regarding the retrieval of data deposited in a given data hub)
- metrics for data deposition (measurements regarding the submmission of data from the data collecotors to a given data hub)
- audit of data requesters' actions (methods the data hubs use to collect information regarding what the users do in their platform)
- audit of authorizations (methods the data hubs use to collect information regarding the permissions any user has in regards to specific stored data)
- sharing of metrics (methods the data hubs use to share and make public the above mentioned metrics)

We provided the chance to answer with multiple selectable options to make the final analysis cleaner. Each question had an "other" alternative, where respondents could specify their answers with open text, if not included in the given choices.

24 institutions answered the survey. The first step in the analysis has been the determinations of the answers to be taken into consideration. Unfortunately, 6 answers were incomplete due to a technical error in the survey platform which temporarily impaired the collection of some of the entries. The entities who encountered the error have been re-contacted but they haven't re-submitted the

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

HEALTHYCLOUD
Health Research & Innovation Cloud

survey. Thus 18 answers were recorded for the data usage metrics section. Out of those, 2 responses have been discarded; the exclusion criteria for both of them has been a self-declaration of the respondee in the comment section of the survey, stating that their data hub is not operational yet and they are not able to provide answers. Thus, a total of 16 pertinent and significant answers has been taken into account for the following analysis.

Given the questions were structured with selectable choices, the answers were counted and presented as percentages. The options added by the respondee have been taken into consideration and discussed in the results chapter, but not quantified.

2.3. Incentive systems

In order to understand the existing incentive systems available and provide a proposal on incentive systems to maximise data sharing and reuse in the context of HealthyCloud, the methodology consisted in three steps:

1. Review the literature to understand the incentive mechanisms that have been used over the past years in the life sciences and beyond.
2. Send a survey to data hubs to understand which data hubs usage metrics they use and what are the incentives systems that are applied in their infrastructures.
3. Gather feedback from the user communities about the incentives systems proposed.

The incentive system proposal was developed in an iterative manner with contribution by experts involved in WP4. To facilitate this process, dedicated workshops were arranged to discuss, evaluate, and enhance the outcomes of the discussions based on the survey responses.

2.3.1. Literature review

In order to avoid reinventing the wheel and designing a new systematic literature review from scratch, secondary studies already published in the literature were searched. It was found that two literature reviews about incentive systems in research were already available.

- Rowhani-Farid et al. (2017) [4] report a literature review that followed the research question used to design the review is "What incentives increase data sharing in the health and medical research community?".
- Devriendt et al. (2021) [5] report the results of systematically reviewing the literature to know "which incentives exist and have been proposed,
  and to critically assess their advantages and disadvantages".

Both studies were aligned with the goal of understanding the incentive systems reported in the literature, making it not necessary to perform a new systematic literature review as part of this task. Aside from those, other papers related to incentive systems were analysed:

- Holub et al. (2018) [6] include some principles that incentives schemes must implement in order to be successful.
- Ali-Khan et al. (2017) [7] include an analysis of motivating participation in open science by examining researcher incentives.
- Eger et al. (2015) [8] include the determinants of open access publishing.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

- Cambon-Thomsen et al. (2011) [9] include the role of a bioresource research impact factor as an incentive to share human bioresources.
- Kidwell et al. (2016) [10] include the importance of using badges to acknowledge open practices as an incentive.
- Neylon et al. (2017) [11] include the role of funders in improving data management and sharing practice amongst researchers.

2.3.2. Survey questions

To understand the data hubs' used incentive systems, 4 questions were included in the survey related to this for the data hubs to answer. The questions were:

- Do you use any incentives to promote the submission of data to your facility?
- Do you use any incentives to promote the reuse of data hosted in your facility?
- Do you use any other type of incentives?
- Do you have knowledge about any other incentives that can be useful for data submission or reuse?

To implement each question, an initial "Yes/No" question was asked and then the open free text ("Describe them") was displayed if the answer was "Yes".

24 entities answered the survey. As mentioned above, out of those, 2 responses were discarded; the exclusion criteria for both of them was a self-declaration of the respondee in the comment section of the survey, stating that their data hub is not operational yet and they are not able to provide answers. The aforementioned technical issue did not affect the incentive part of the survey. Thus, a total of 22 pertinent and significant answers were taken into account for the analysis.

2.3.3. Stakeholders' feedback

The WP8 is organizing a series of workshops with external stakeholders of interest. To validate the proposed incentive system, a dedicated session in one of the workshops was set (HealthyCloud user communities, March 2023).

The workshop had 20 participants (experts in the field, but not from the data hubs surveyed before) and the incentives were discussed on the second day in a series of round tables. To drive the discussions, a small form was sent to them previously to the session, so that they could rate the proposed incentives (taken from the literature review and survey sent to data hubs) depending on the impact they believe they have as well as their ease of implementation. Then, during the round tables, all participants could give feedback about the incentives and mention which ones would be more useful and interesting for them.

After the workshop, the form that had been previously sent to the workshop participants was circulated through the HealthyCloud and ELIXIR newsletters. This was done to gather additional responses and gain a clearer understanding of the researchers' opinions on the proposed incentives. A total of 25 answers (including the stakeholders' workshops participants) were gathered.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

3. Results

3.1. Auditing of data hubs' usage

3.1.1. Survey collected answers

As stated above, we have collected answers to 5 questions regarding different aspects of data usage auditing and metrics. In each question, the respondents could select as many answers as they needed. In here, we depict the distribution of the answers:

Table 1. Raw answers to the question regarding metrics of data usage, excluding comments added as free text.

How do you measure data usage in your data infrastructure?		
multiple choice answers	**percentage**	**# of selections / 16**
number of requests	87,5 %	14
number of requesters	62,5 %	10
country of origin of the request	62,5 %	10
data downloads and/or access	62,5 %	10
date of data downloads and/or access	31,25 %	5
we do not measure data usage in any way	6,25 %	1

Table 2. Raw answers to the question regarding metrics of data deposition, excluding comments added as free text.

How do you measure data deposition in your data infrastructure?		
multiple choice answers	**percentage**	**# of selections / 16**
number of deposited datasets	75 %	12
number of data controllers (DACs)	43,75 %	7
country of origin of the data	18,75 %	3
date of data deposition	50 %	8
we do not measure data usage in any way	6,25 %	1

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Table 3. Raw answers to the question regarding audit of data requesters actions, excluding comments added as free text.

How do you audit data requesters' actions?		
multiple choice answers	**percentage**	**# of selections / 16**
record of entry in the data infrastructure	56,25 %	9
record of data request	62,5 %	10
record of data download or access	56,25 %	9
we do not audit data requesters' action in any way	12,5 %	2

Table 4. Raw answers to the question regarding audit of authorizations, excluding comments added as free text.
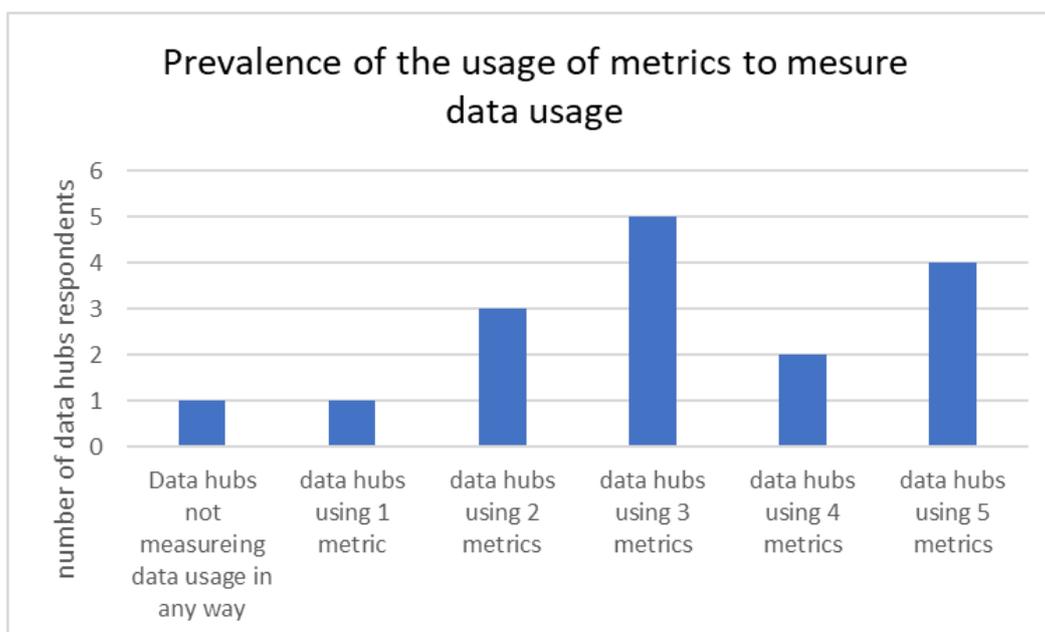
How do you audit data access authorizations?		
multiple choice answers	**percentage**	**# of selections / 16**
record of authorization changes	43,75 %	7
record of changes in group or role memberships	50 %	8
we do not audit data access authorization in any way	18,75 %	3

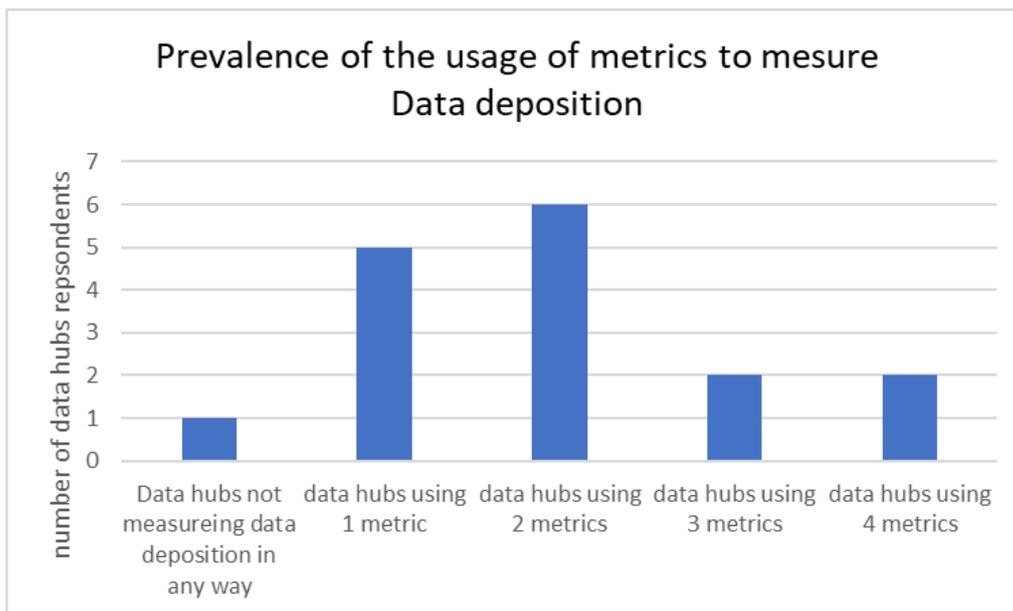**Table 5**. Raw answers to the question regarding sharing collected metrics, excluding comments added as free text.

How do you make public the metrics regarding data deposition and usage?		
multiple choice answers	**percentage**	**# of selections / 16**
showing them in the infrastructure website	50 %	8
actively sharing the information with patients and or citizens	6,25 %	1
actively sharing the information with funders or any other institution	37,5 %	6
actively sharing the information with data controllers	12,5 %	2
we can share them with proper stakeholders under specific request	37,5 %	6
we do not make available any metric	12,5 %	2

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

3.1.2. Analysis of the results

Metrics for data usage and deposition. From the answers related to data usage and data deposition metrics, we can see that the use and record of metrics is a very common practice in the current European health data ecosystem, as shown in figures 1 and 2. we see that 15 out of 16 respondents declared to use at least one metric for both data usage and data deposition. In addition, we learn that the vast majority of Data hubs use several (more than 2) metrics to measure and record data usage (87,5 %) and data deposition (75%). The number of used measures are depicted in figures 1 and 2. Number of requests and number of depositions are the most common measures, consistent with them being the easiest to capture. Nonetheless, more than half of the respondents declared the use of additional metrics like country and date of requests and depositions. In addition to the provided selectable answers, regarding data usage, some respondents specified that they measure the types of data being requested, types of requester (e.g. academic/commercial, ECR etc), protocol and speed of data download. Others specified that they record metadata views, output projects and publications; these last ones are not strictly metrics of data usage, but related secondary measures. Regarding data deposition, two respondents explained that they measure volume of data and population coverage instead of number of datasets.
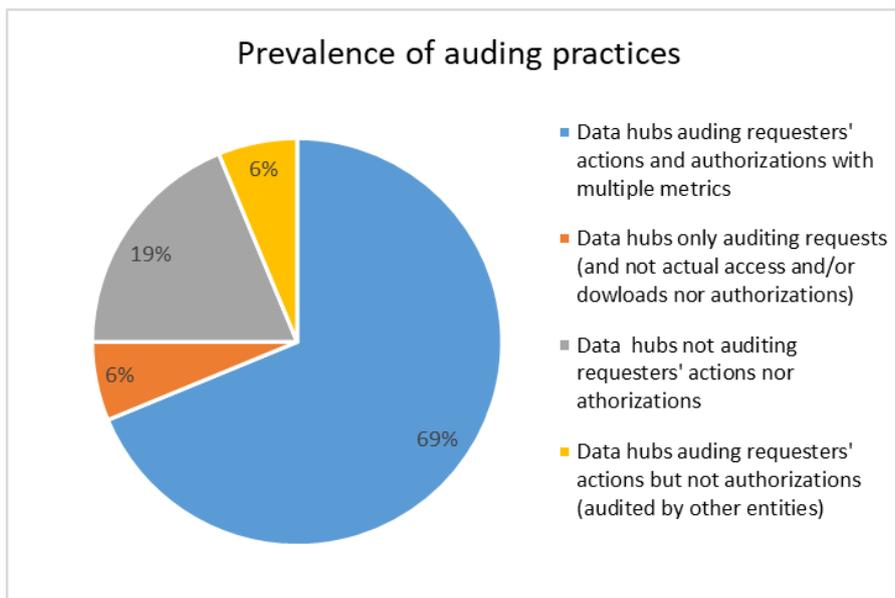

Figure 1. Graphical view of aggregated results from data in table 1.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Figure 2. Graphical view of aggregated results from data in table 2.

From the answers regarding the auditing of requesters' actions and data access authorizations, we can infer that such auditing practices are highly prevalent among European Data hubs. As illustrated in figure 3, the majority of the data hubs perform auditing of the actions carried out in their platform. Three respondents declared to not audit actions nor authorizations in any way. One of those three, declared so because their Data hub is in the planning phase and the auditing strategy has not been decided yet. One data hub declared to audit uniquely data requests, but not actual data access or download; at the same time they declared to not audit authorizations either. Another data hub working as an aggregate of biobanks and data collections described that they audit all data requesters' actions but do not handle nor audit authorizations. In addition, several respondents declared auditing/recording of actions and other aspects going beyond data access/download such as:

- patient and public engagement related to the data access request
- publications and other output related to the data
- use of the data adhering to commitments to data security
- Applications for data use
- correct use of data, remote checked using an annual report that the data users have to submit
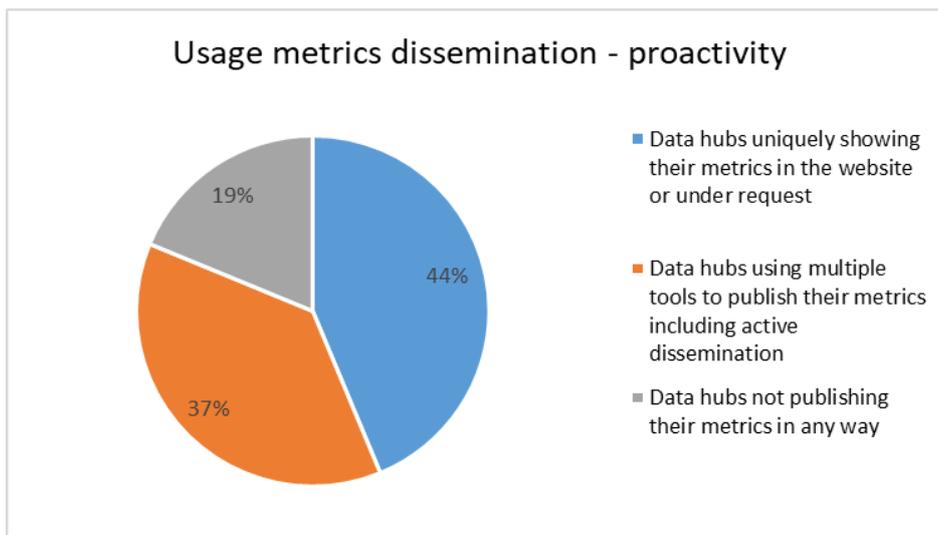
D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Figure 3. Graphical view of aggregated results from data in tables 3 and 4.

Finally, we inquired whether and how the Data hubs disseminate the metrics relative to their usage. This would include metrics regarding data deposition as well as data retrieval. Collectively, this practice appears quite common, with only 3 data hubs declaring not to share their metrics at all. The most widely used tool is the own data hub website, where 50% of the respondents declare to show their metrics. Just below 40% of the respondents stated that they actively share the usage metrics with their funders or other institutions and the same amount (not complete overlap) declare that they can share them with stakeholders under request. Only one respondent shares the metrics with data donors so far, but an extra one declared in the comments to be planning to do so as well. A transversal analysis of the answers illustrated in figure 4 reveals three levels of engagement and proactivity in relation to this topic:
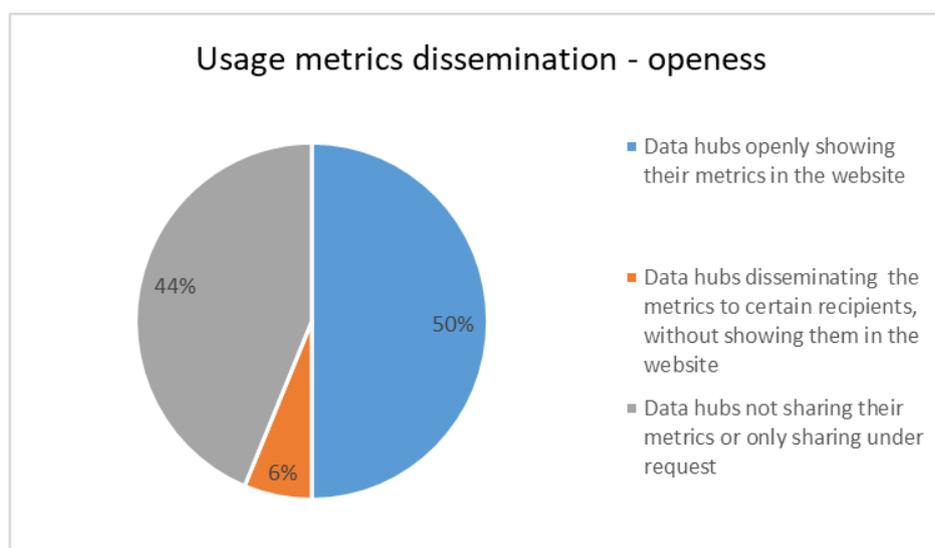
- Below 20% of the respondents do not make public nor available any usage metric.
- Above 40% of the respondents have their usage metrics available in their website or either they have them stored internally and can share them under specific request.
- Below 40% of the respondents use multiple strategies to share their usage metrics, including actively disseminating to funders/citizens/DACs (see broken numbers in table 5).

In parallel, the same answers can be stratified according to the level of openness in relation to this topic:

- Above 40% of the data hubs do not share their metrics or only share them under specific requests.
- 6% of the data hubs disseminate the metrics to certain recipients, without showing them in the website.
- 50% of data hubs publish their usage metrics openly on their website.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Figure 4. Graphical view of aggregated results from data in table 5.

Figure 5. Graphical view of aggregated results from data in table 5.

3.2. Incentive systems

3.2.1. Literature review

Medical research is unique compared to other scientific fields because access to medical data relies on voluntary contributions from individuals such as patients or doctors. To encourage data sharing and promote reuse, a positive incentive system should be established and embraced by the research community. These incentives can encourage resource providers to implement transparent access policies and simplify access procedures. For these incentives to be successful, MartinGillian et al. (2018) [6] present some principles that the incentives should follow:

1. Incentives must be in place for all the links in the chain: (1) biological material and data generation or collection; (2) biological material and data storage, curation, and enrichment; and (3) biological material and data reuse.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

2. For biological material and data collections receiving public funding or infrastructural funding, the incentive must stimulate reuse by external users, namely users outside the infrastructure.

3. Contributions to existing biological material and data collections should be supported by funding organizations.

4. Academic promotion schemes and institutional evaluation schemes should incorporate contributing to and reuse of existing biological material and data collections.

In addition, Holub et al. (2018) [6] also mentions that to encourage the reuse of biological materials and data, it is essential to establish incentives for both the providers and the users. Researchers should have a motivation to use existing resources, while resource providers must also have clear incentives to facilitate the reuse of their materials. To enforce these principles, funding agencies, academic organizations and publishers should apply pressure on resource providers to implement transparent access policies and make their resources more easily accessible to demonstrate their reuse.

Many research infrastructures apply incentive systems that follow the previous principles and that support research data sharing and reuse. To understand the systems that are used nowadays in a European context, a literature review (see Methodology section) was done. A common understanding of the review done is that many authors identify credit as a barrier towards sharing data, so that incorporating attribution is crucial. Finally, to better understand what incentives are useful for the wider community, it is also important to understand the motivations of researchers to participate in open science. For that, Ali-Khan et al. (2017) [7] studies these motivations and they are considered when defining the incentives. Having all these in mind, table 1 includes the proposed incentive systems.

Table 6. Incentives to promote data sharing taken from the literature.

Incentive	Description
Co-authorship	Data generators' inclusion as co-authors in return for sharing their research data with secondary users.
Dataset attribution	Unambiguous linkage of the data to their contributors via a unique identifier to acknowledge individual endeavor and achievement. Data citation can be achieved by submitting to repositories that automatically attach data DOIs, so the dataset is linked to the list of investigators involved in collecting the data.
Dataset citation	Recognize published datasets as a citable entity (publication).
Data advertisement	Data collected by the data partners in a consortium is advertised, as well as the identity of the data partners, including their logos and URL links. This provides opportunities for groups to expand their public visibility while retaining control of their data, and thus increase the use of their datasets and subsequent citations.
Dataset citation and usage statistics	Use dataset citations and usage statistics and other user feedback to be used as important measures of credit. This would include development of new metrics to measure sharing contributions (so that they could be included in funding and academic advancement decisions).
Ethical and legal framework	These include the belief that publicly-funded research outputs ought to be released with minimal delay and that consistent with patients' wishes, researchers have a duty to ensure that samples are broadly shared to maximize research and discovery. For that, an ethical and legal framework to facilitate a culture of responsible and effective sharing of sensitive data would be needed.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

HEALTHYCLOUD
Health Research & Innovation Cloud

Data sharing policy	Implementation of a data sharing policy that aligns the reward system to ensure that scientists sharing data are acknowledged/cited and that data sharing is credited in research assessment exercises and grant career reviews.
Embargoes	An embargo in this context means holding a monopoly on publications from a certain dataset within a fixed time period. Embargoes are a way to deal with researchers' concerns that rapid data sharing might lead to other groups publishing important results before them. As such, embargoes are limited in both time and scope, and essentially safeguard the 'right of first publishing' of the data generators.
Collaborations	Sharing data will build new collaborations and spur the development of new tools and technologies.
Career advancement	The potential for data sharing to bolster the professional standing of researchers. Requests for reagents, data or other tools leads to diverse and unanticipated collaborations, expanding researchers' interests, visibility and professional impact. Also, sharing has a positive impact on publication productivity and citations.
Institutional support	Institutional support is critical to support data sharing amongst researchers since making data ready is time and resource-intensive, including the need for payment of open access publication fees, and the preparation, formatting and handling of data and other research outputs to make it ready for sharing.
Public and patient benefit	Sharing research outputs would increase patient and public trust, increasing transparency and channeling benefits to patients, rather than to industry. This could motivate high-quality researchers and trainees and increase patient participation in research. Also, greater transparency through the research process would allow patients to become more equal and informed research partners. Improvement of researcher-participant relationships could enable more patient-responsive studies, enhancing patient satisfaction and research outcomes.
Intellectual property rights	Intellectual property rights over discoveries derived from researchers' data.
Electronic lab notebooks	Introduction of an electronic lab notebook that allows the deposition of all primary data as well as data management and coordination tools that enhance community input in an easy way.
Badges for open data	Badges to acknowledge open data, open materials and preregistration of research if published. The criteria to earn an open data or open materials badge should involve making all digitally shareable data and materials available in an open data repository.
Harmonization of ethical requirements	Harmonizing ethical requirements internationally is a key priority to realize the promise of global open science. One way that harmonization could accelerate open science would be the introduction of standard click-wrap agreements that are signed when patient-derived data are downloaded, thus avoiding the need to use traditional (and cumbersome) side agreements that are signed (and sometimes notarized) by central administrators.

Considering the table above, it is possible to see that some of the incentives are related. For instance, acknowledgement and citation are included in 4 different incentives (co-authorship, dataset attribution, dataset citation and dataset citation and usage statistics). These incentives are important for building trust and accountability in the scientific community. By providing proper credit and recognition, researchers can build a culture of trust and collaboration that encourages the responsible sharing and reuse of data.
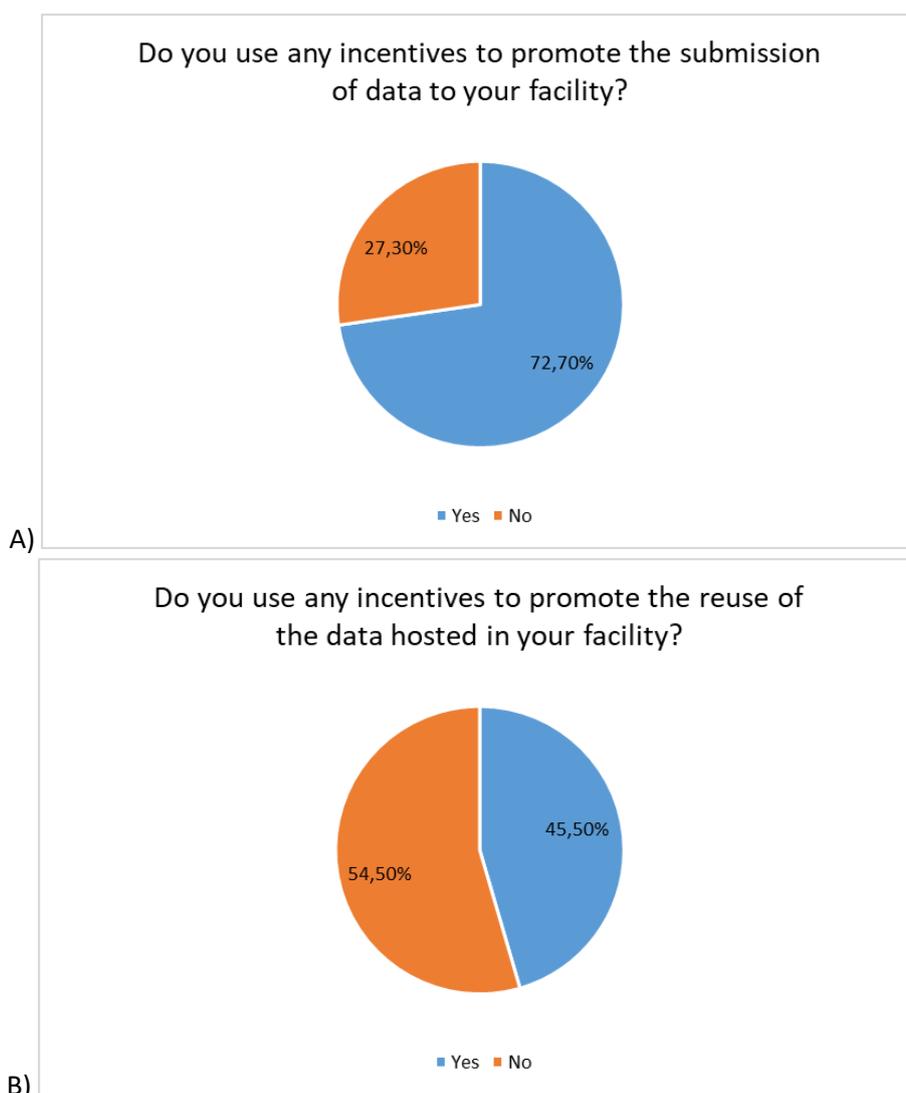
Other incentives include policy and regulation requirements, that can be powerful drivers for promoting data sharing and ensuring adherence to data standards. Such incentives can include requirements for data submission to a health data hub, legal obligations for sharing data, and

D4.3 Report on the usage and auditing of the data hubs indicators
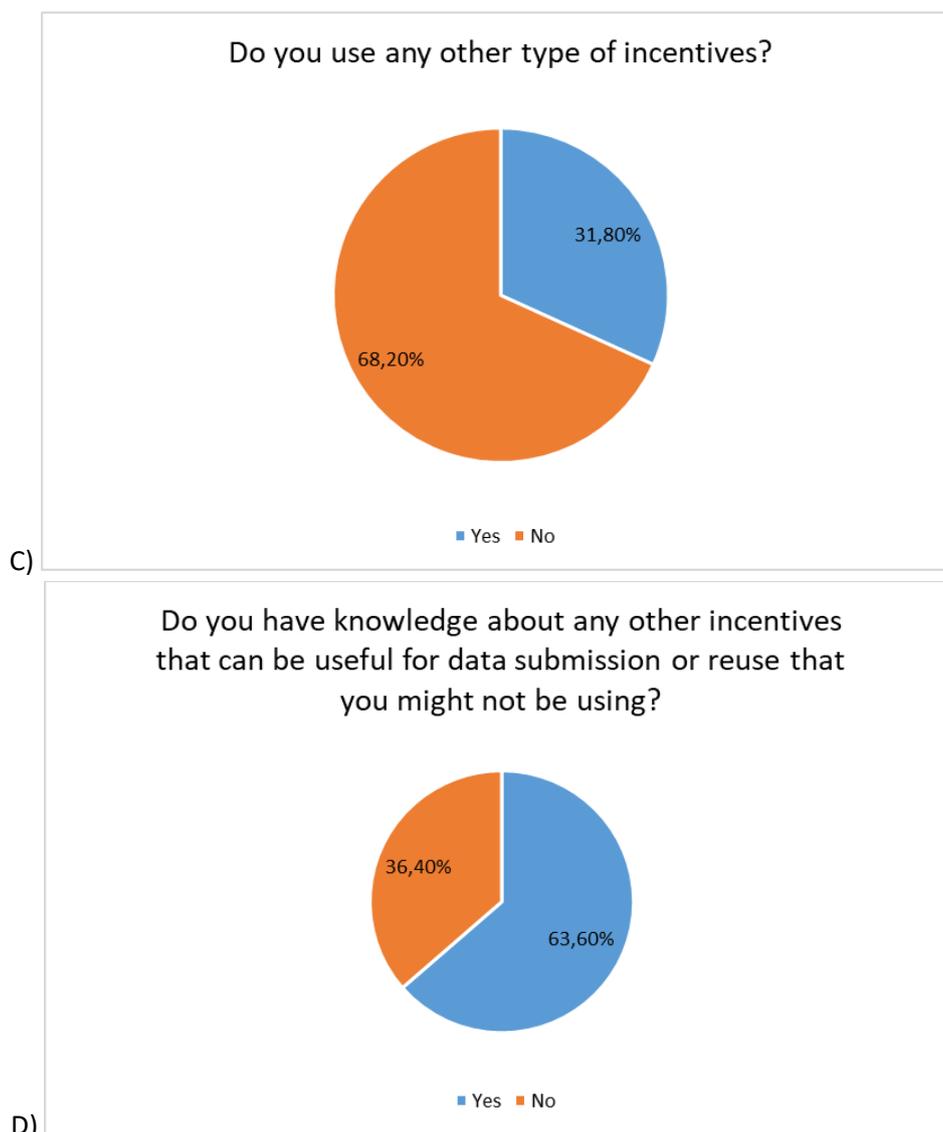and proposal of an incentive system
Version 2.0

mandating data sharing as a prerequisite for publication. By establishing clear policies and requirements, health data hubs can not only enhance the quality of data shared and improve data accessibility, but also improve the legal certainty surrounding data sharing. This increased clarity and transparency can ultimately advance scientific research and healthcare outcomes, benefiting both researchers and citizens [12]. This includes the incentives on ethical and legal framework, data sharing policy, institutional support, and harmonization of ethical requirements.

Finally, collaborations and data advertisement can play an important role in promoting data sharing and improving the impact of health research. These incentives, can help increase awareness and encourage more researchers and institutions to participate in data sharing. Related to this, patient incentives, such as transparency about the research project performed based on their EHR, can help build trust and promote patient engagement in health research.

3.2.2. Survey collected answers

As mentioned in the Methodology section, 4 questions about incentives were asked in the survey to data hubs. All the questions were "Yes/No" followed by an open free text ("Describe them") if the answer was "Yes".

A)

B)

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

HEALTHYCLOUD
Health Research & Innovation Cloud

Figure 6. Survey answers of the incentives questions.

The answers of the free text are found in annex III.

Before the analysis of the results related to the incentives mechanisms, it was necessary to identify the systems proposed that were actual incentives. To do that, the answers were divided between two categories: incentives and benefits. This was done because some of the answers were considered as intrinsic rewards or advantages that are received for being a data hub, and they were not any particular action that a data hub does to increase the number of submissions or the reuse of their data. The definitions used for this classification were the following:

- Incentive: a stimulus that motivates or encourages a person to take a particular action. It is designed to influence behavior or performance. In this case, an incentive would be an extra action that a data hub should do to increase submission/reuse.
- Benefit: a reward or advantage that is received as a result of some action or circumstance. In this case, benefits are intrinsic to the way a data hub works.

After that, the next step was to match the incentives that were reported by different data hubs. This

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

way, it was possible to see if there is a system that is applied by more than one data hub and therefore might be a good one to consider for the project. Once this was done, the incentive systems that were answered in the 'others' question were also classified between submission or reuse incentives. Finally, some of the incentives were taken out of the classification as they were not incentives for the data hubs to do, but incentives that come from policy makers, funders or publishers. However, these incentives are very important and will be taken into account for the final proposition of the incentive system. In summary, 18 incentives were considered to promote submission of the data and 6 to promote the reuse of the hosted data. Following tables (from 7 to 10) includes the final classification.

Table 7. Incentives used by data hubs to promote submission of data.

Incentives used by data hubs to promote submission of data	# of data hubs
Implement a reimbursement model	3
Give pseudonymization and de-identification services	2
Return the results as part of the access policy	1
Disseminate the submitted data, providing outreach opportunities	1
Introduce and give support in regulation, guidelines, standards and new policies	1
Give training and support in data quality and curation	1
Promise feedback about the information taken from the data	1
Maintenance and legal ground for future uses	1
Provide quality improvement reports	1
Provide data stewardship as a service, crediting data stewards for supporting submissions hence promoting submission of quality datasets	1
Host data shared for scientific research free of charge	1
Harvest metadata of datasets to other services to increase discoverability (e.g., OpenAire Explorer)	1
Discount access to services in return to bring your data	1
Advertise what good has come out of submitting data	1
Provide long-term archiving for a fixed/low price	1
Exclusivity of use for some partners for some time before opening up the novel data to all researchers	1
Converting data to a common schema (e.g. OMOP or FHIR) to increase interoperability	1

Table 8. Incentives used by data hubs to promote reuse of data.

Incentives used by data hubs to promote reuse of data	# of data hubs
Reduce access fees for low and middle income countries	1
Provide easy request access procedures (saving time)	1
Make packages free of charge if third party usage is allowed	1
Cite hosted datasets in papers as scientific publications	1

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Highlight exemplar data returns and/or use of the resource, which can be helpful to researchers' career development/visibility	1
Showcase data to improve reuse and thus citation of the publications	1

Table 9. Benefits to submit data to a data hub.

Benefits to submit data to a data hub	# of data hubs
Increases findability (metadata)	3
Provides data hosting during a research project, which is mandatory for european calls	2
Increases visibility	1
Ensures sustainability upon project end	1
Ensures recognition from the research community	1
Promotes networking (scientific collaborations)	1
Provides knowledge and expertise of partners, consultancy, existing governance arrangements	1
Ensure data quality and curation	1
Cover costs of compute and data storage	1

Table 10. Benefits to reuse data from a data hub.

Benefits to reuse data from a data hub	# of data hubs
Makes the resource highly accessible	1
Promotes national health database	1
Promotes networking (scientific collaborations)	1
Makes collaborators visible by adding information of the data and the team that produced it in a common portal	1

The incentives that were taken out the classification as they were out of the reach of a data hub, were the following:

Data submission
- Make it a requirement to receive funds.
- Make it a requirement by publishers / journals.
- Make it a legal obligation.
- Make data submission to a data hub part of a deliverable.

Data reuse
- Highlight it as best practice.
- Fund research projects that reuse data.
- Give extra funding to research projects that make data accessible for reuse.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0
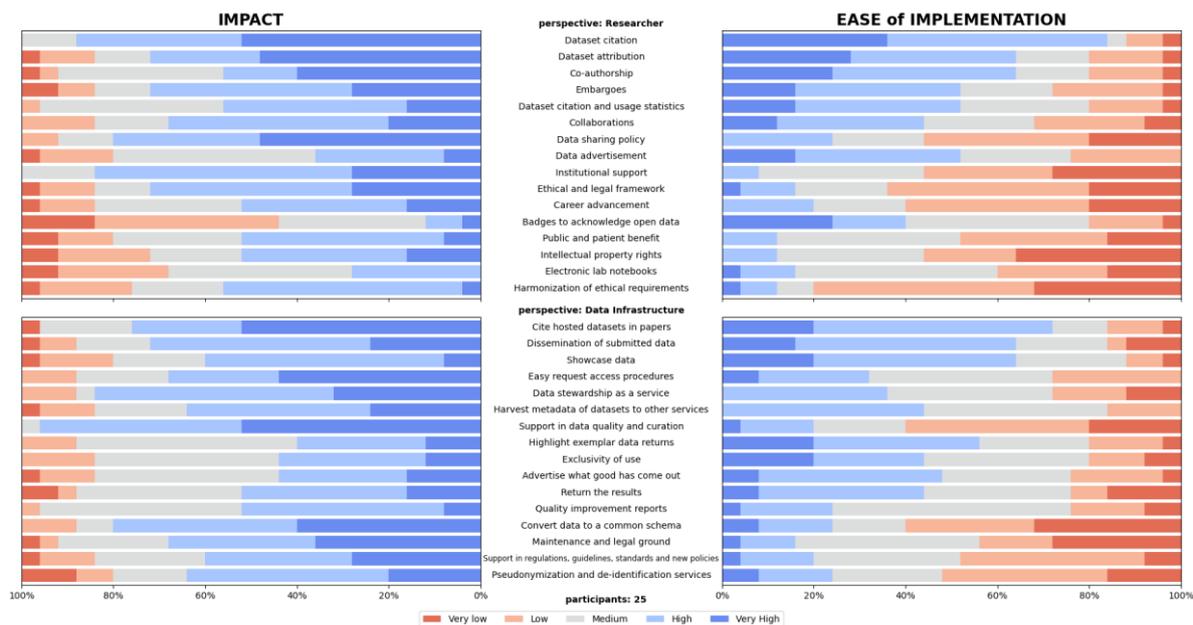
3.2.3. Stakeholders feedback

During the stakeholder workshop, the participants, that were part of the user's community that the HRIC would have, mentioned that the most interesting incentives for them would be the following:

- Create a **usage and citation index** to understand how much the dataset is cited, so that other organizations can use it. This index can be then used as a measure to fund distribution and as an impact factor, since if a dataset is highly cited it should mean that it is a highly curated dataset with good quality.
- **Showcase** somehow the good impact of sharing the data.
- **Economic incentives** to get the resources to be able to curate and prepare the data to be shared, as sometimes even though they would like to share they do not have the resources to do it.
- Have a **good data strategy** from the beginning and create an added value from the data generated and use this to build collaborations with other entities that also work like this.
- **Data sharing policy** mandatory for institutions, so that researchers feel safe when sharing their data, as right now most of the time feel that their institutions do not back them up when they want to share their data. This policy should be easy to understand and transparent so that it is a way to help researchers to share their data.

During the discussion it was also mentioned the importance of raising awareness to get the mind shift. It is necessary for journals and conferences to promote data sharing as many of them still do not require the data to be shared to publish, and it is possible to see a clear bias for the journals that do not require data to be shared. As publishing is the main priority to most researchers, it influences that many of them use their time to get more publications out and then do not have time to make their data FAIR and ready to be shared.

As mentioned in section 2.3.3, the form that the participants of the workshop used to drive the discussions was later sent to the HealthyCloud and EOSC newsletters. The results from this form are shown in Figure 7.

Figure 7 illustrates two perspectives: the researcher's viewpoint and the data hubs' viewpoint. The top section represents the researcher's perspective, where incentives derived from the literature review are listed and rated. Conversely, the bottom section represents the data hubs' perspective, incorporating incentives gathered from a survey sent to them. Additionally, the figure uses color coding to highlight specific indicators. The blue indicators represent the responses from the survey participants who believed these indicators would have the greatest impact and would be relatively easier to implement. Conversely, the red indicators represent the responses indicating the least impact and perceived difficulties in implementation according to the survey respondents.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Figure 7. Results from the form sent to HealthyCloud stakeholders' workshop participants and HealthyCloud and EOSC newsletters, to identify the opinion of the users on the impact and ease of implementation of the incentives.

3.2.4. Analysis of the results

To evaluate the effectiveness of the proposed incentive systems and determine the most suitable one for the project, a series of workshops were conducted with the members of WP4. The main outputs of these workshops were the following:

- When it comes to incentives for data submission, providing training and support on regulations, new policies, standards, and the data submission process to the data hubs was considered highly beneficial. By establishing clear guidelines and standards for data sharing, data can be more easily shared and reused, leading to more meaningful research outcomes. Additionally, providing training and support on these regulations and standards can help ensure that data is collected and shared in a consistent and standardized manner, ultimately improving data quality and reliability. This aligns with the concept of offering data stewardship as a service, which was also regarded as important in facilitating data submission by users.

- To incentivize data reuse, showcasing data to improve its reusability was seen as an excellent approach that could be easily implemented by the data hub. Additionally, it was highlighted as significant to cite hosted datasets in research papers, as acknowledgements serve as motivation for researchers to publish their data. These citations should highlight how the provided data has contributed to positive outcomes.

- Among the benefits, ensuring data quality and curation was considered of utmost importance, as researchers take pride in being associated with reputable data hubs, thereby enhancing the recognition of the data hub itself. Furthermore, presenting information about the data hubs within the same portal was considered useful if the information is well-structured to avoid confusing users while providing them with necessary details. Lastly, promoting national health databases was considered challenging, especially in countries where such databases do not exist uniformly across all hospitals, as each may maintain its own independent database.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

- The discussions in the workshops also revealed that many of the incentives revolved around monetary rewards. Specifically, five of the proposed incentives aimed to reduce costs or provide free access for certain actions. However, it was recognized that the incentive system for HealthyCloud should not primarily rely on monetary considerations, as hosting data should ideally be free. Additionally, implementing a reimbursement model might raise ethical concerns.
- Providing feedback about the information extracted from the data was identified as a challenging and time-consuming task for data hubs. However, participants suggested that a valuable incentive in the same vein would be to provide feedback on the data quality itself.

Comparing the results from the literature review and the survey, several common themes emerge. This comparison enable us to identify what incentives are interesting for researchers the data hubs are already using or considering for implementation.

- Both the literature review and survey emphasize, dataset attribution and citation, as well as data advertisement. Recognizing and crediting datasets used in research is crucial, and promoting shared data increases its impact and attracts users.
- The establishment of an ethical and legal framework, along with a data sharing policy, are also essential incentives for researchers and data hubs. This includes giving support on regulation, guidelines, standards, and new policies.
- Career advancement, which entails gaining recognition from the research community through data citations, embargoes granting exclusive use of novel data to select partners before opening it to all researchers, and promoting networking and scientific partnerships are consistently highlighted as significant incentives for both data hubs and researchers.

During the HealthyCloud stakeholders' workshop, participants highlighted the importance of creating usage and citation indices to measure the impact of shared data and offering economic incentives to support data preparation and curation. Participants also expressed a desire for a mandatory data sharing policy for institutions that is easy to understand and transparent, to help researchers feel safe when sharing their data.

Based on the form results, it is evident that once again, incentives related to citation and attribution are rated as having a higher impact and being relatively easy to implement. Conversely, many indicators are deemed to have a high impact but pose challenges in implementation. These include harmonizing ethical requirements, which is a complex task across different countries, establishing an ethical and legal framework as well as harmonization or ethical requirements, both of which researchers often struggle with, and providing support for data quality and curation, that is recognized as a tedious and time-consuming endeavour, as is the conversion of data to a common schema. On the other hand, incentives such as using electronic lab notes or badges to acknowledge open data are perceived to have a low impact according to the ratings.

In summary, the importance of citation and the establishment of a data usage and citation index emerged consistently across all three cases: the literature review, the survey, and the stakeholders' feedback. Promoting data through advertisement and highlighting the impact of data sharing were also mentioned in all three instances. However, it is essential to increase researchers' awareness about the advantages of data sharing in order to facilitate a change in mindset. While data hubs can provide support in terms of new regulations and policies, it is crucial to recognize that certain incentives, such as data sharing policies, ethical and legal frameworks, and institutional support, require actions beyond the scope of data hubs' capabilities. Instead, data hubs can assist in advocating for and facilitating the implementation of these incentives when necessary.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

4. Conclusions

4.1 Auditing of data hubs' usage

In today's rapidly evolving technological landscape, the utilization of data hubs has become increasingly common. However, with this rise in usage comes the need for effective data management , metrics for evaluation and incentive systems to ensure continued growth and sustainable success.

Surveys have emerged as a critical tool for analyzing metrics and incentive systems utilized within data hubs. Beyond gathering quantitative data, surveys can also provide valuable qualitative insights that help identify challenges and key actions necessary for the effective development of these infrastructures. By targeting existing data hubs, surveys allow researchers and industry leaders to gain a deep understanding of how these systems are being used in practice and what measures can be taken to optimize their performance.

From the analysis of the results exposed above we can conclude that measuring data usage and deposition in their infrastructure is a common practice among data hubs. The vast majority (15 out of 16) of the data hubs answering our survey declared to measure their usage. It is quite predictable that such infrastructures have a measurable quantification for their usage, thus being able to evaluate their work and activity. Some metrics are much more common than others, like number of deposited datasets and number of requests.

> ➢ The new HRIC should recommend the use of a few common metrics for the included data hubs, thus giving the possibility to collectively measure the total volume of data usage.

It is important to take into account in the strategic agenda that the HRIC portal must allow the incorporated Data hubs to keep measuring their actions as they do at the present, even when data deposition and usage is channelled through a common portal. An impairment of this would decrease the hubs' willingness to join it, thus seriously hampering the development of a comprehensive HRIC.

As well as measuring usage, also auditing users' action is a common practice among the respondent data hubs. The variability in this aspect exposed in the results is mostly due to the different levels of delegation and aggregation of the infrastructures: some data hubs expose data actually hosted in other smaller infrastructure, thus they do not measure the authorisations and access.

> ➢ The new HRIC shall recommend the use of some common auditing systems which virtually all data hubs could implement. In this way we could establish a minimal level of tracking of actions to harmonize the various participating data hubs.

Once we captured and exposed how and how much the data hubs measure their work, we asked how they disseminate the results. The distribution of the answers in this aspect is scattered and they show various levels of proactivity and openness (see results section 3.1).

> ➢ The future HRIC should advocate for active engagement in sharing usage metrics, perhaps centralizing the collection of information to be displayed in the HRIS portal as the first step.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

The dissemination of the data regarding usage, legal compliance, user experience and benefits to the scientific advances is a crucial step in the dialogue with the relevant stakeholders, including funding agencies.

4.2. Incentive systems

After conducting a thorough literature review, survey, and stakeholders' feedback, it is evident that there are various incentives that could encourage data generators to share their data and data users to reuse data that is available in data hubs.

Citation and the creation of a data usage and citation index were consistently identified as crucial factors in all three cases: the literature review, the survey, and the stakeholders' feedback. Additionally, promoting data through advertisements and emphasizing the impact of data sharing were mentioned across all three cases.

It is worth noting that monetary incentives may not be ethical or effective in encouraging data submission and reuse. In addition, some incentives are considered highly impactful but face implementation challenges. For instance, having an ethical and legal framework and harmonizing ethical requirements across different countries are complex tasks, and providing support for data quality and curation can be a time-consuming endeavour, as is the conversion of data to a common schema.

It is important to emphasize the need to increase researchers' awareness of the benefits of data sharing to foster a change in mindset. Additionally, many incentives that could be implemented are beyond the control of data hubs. Incentives such as data sharing policies at research institutions, requirements for journals to publish data, and securing funds for projects that generate or reuse data are all factors that can increase data submission and reuse. However, the responsibility for implementing these incentives lies with policy makers, publishers, and funders, rather than solely with data hubs. While data hubs can provide support in terms of new regulations and policies, it is crucial to recognize that certain incentive, require actions beyond the scope of data hubs' capabilities.

5. Next steps

The HealthyCloud project has a logical workflow in which several pieces of information on diverse aspects are gathered by the different Work Packages and the conclusions are taken into consideration for their incorporation, direct or indirect, in the final product. The conclusions from this deliverable follows precisely this concept, and will be read, actively communicated, and thoroughly analyzed by WP8, in charge of delineating the strategic agenda.

In addition to this, partners of this WP will use the material collected, especially from the deep analysis of the incentive systems landscape, to write an opinion paper to a scientific open access journal. In that way, we can put our analysis and expertise at the benefit of the scientific community at a bigger scale. We consider that the major agents of incentive systems indeed, should not be the Data hubs themselves, who, already offering a free of charge service are in a hard position to increase their engagement, but rather funding agencies and journal editors. They have the means to actually enforce the deployment of correct Data management practice as Data deposition, sharing and reuse of existing data.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

6. References

[1] Glossary of commonly used terms in the field of health data research - developed by the EU project HealthyCloud: https://doi.org/10.5281/zenodo.5997584

[2] Yaqoob, I., Salah, K., Jayaraman, R. *et al.* Blockchain for healthcare data management: opportunities, challenges, and future recommendations. *Neural Comput & Applic* 34, 11475–11490 (2022). https://doi.org/10.1007/s00521-020-05519-w

[3] National Institute of Standards and Technology (NIST). Metrics. c2022 [cited 2022 Feb 23]. Available from: https://csrc.nist.gov/glossary/term/metrics

[4] Rowhani-Farid, A., Allen, M. & Barnett, A.G. What incentives increase data sharing in health and medical research? A systematic review. *Res Integr Peer Rev* 2, 4 (2017).

[5] Thijs Devriendt, Mahsa Shabani, and Pascal Borry.Data Sharing in Biomedical Sciences: A Systematic Review of Incentives. *Biopreservation and Biobanking*. Jun 2021.219-227.

[6] Holub P, Kohlmayer F, Prasser F, Mayrhofer MT, Schlünder I, Martin GM, Casati S, Koumakis L, Wutte A, Kozera Ł, Strapagiel D, Anton G, Zanetti G, Sezerman OU, Mendy M, Valík D, Lavitrano M, Dagher G, Zatloukal K, van Ommen GB, Litton JE. Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. *Biopreserv Biobank*. 2018 Apr;16(2):97-105.

[7] Sarah E Ali-Khan, Liam W Harris, E Richard Gold (2017) Point of view: Motivating participation in open science by examining researcher incentives *eLife* 6:e29319.

[8] Eger, T., Scheufen, M. & Meierrieks, D. The determinants of open access publishing: survey evidence from Germany. *Eur J Law Econ 39*, 475–503 (2015).

[9] the BRIF workshop group. The role of a bioresource research impact factor as an incentive to share human bioresources. *Nat Genet 43,* 503–504 (2011).

[10] Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg LS, Kennett C, Slowik A, Sonnleitner C, Hess-Holden C, Errington TM, Fiedler S, Nosek BA. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLoS Biol.* 2016 May 12;14(5):e1002456.

[11] Neylon, C. (2017). Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers. *Research Ideas and Outcomes*, *3*, e21705.

[12] Cohen, I.G., Amarasingham, R. and Shah, A., 2017. Health care data sharing incentives for aligned performance. Journal of the American Medical Association Internal Medicine, 177(7), pp.966-967.

Annex I – List of contacted Data Hubs

Datahub name	Country	Answered to the survey
National health information system at Gesundheit Österreich GmbH	Austria	YES
Austrian National Covid-19 Data Platform	Austria	YES
Cancer registry	Austria	NO
Cause of death statistics	Austria	NO
UK Biobank	UK	YES
Clinical Practice Research Datalink (CPRD)	UK	YES
Danish Health Data Authority	Denmark	NO
Red Nacional de Biobancos (RNBB) ISCIII	Spain	NO
BIFAP	Spain	NO
BIGAN (IACS)	Spain	NO
EUDAT CDI	Worldwide	YES
European Platform on Rare Disease (EU RD Platform)	Europe	NO
Findata	Finland	NO
THL Biobank (THL - Finnish Institute for Health and Welfare)	Finland	YES
Statistics Finland	Finland	YES
Finnish Social Science Data Archive	Finland	YES
SIB	Switzerland	YES
Ireland - eHealth Ireland	Ireland	NO
Health-RI XNAT BMIA	Netherlands	YES
EGA (CRG - European Bioinformatics Institute EBI)	Europe	YES
DATAMIND: data hub for mental health research	UK	YES
Alleviate	UK	NO
DATA-CAN - Hub for Cancer	UK	NO
Gut Reaction - Hub for Inflammatory Bowel Disease	UK	NO
Discover-NOW - Hub for Real World Evidence	UK	NO
INSIGHT - Hub for Eye Health	UK	YES

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

PIONEER - Hub for Acute Care	UK	NO
BHF Data Science Centre	UK	NO
Estonian Biobank	Estonia	YES
National Health Foundation	Estonia	NO
Medical Informatics Initiative Germany	Germany	YES
DaTraV, Research Data Center	Germany	NO
ELIXIR-LU transmed data hub	Luxembourg - Europe	YES
MMMI- Multi-site Multimodal Molecular Imaging Italian node	Italy - Europe	YES
BCU Imaging Biobank (BCU-IB)	Italy	YES
SDN-Biobank	Italy	YES
BBMRI-ERIC	Europe	YES
EATRIS-ERIC	Europe	YES
DisGeNET	Spain	NO
Norwegian Institute of Public Health	Norway	NO
Andorra health data hub	Andorra	NO
Regionernes kliniske kvalitetsudviklingsprogram (EHR data)	Denmark	NO
EuroBioImaging Med Hub	European	YES
Croatia - Hrvatski Zavod Za Javno Zdravstvo	Croatia	YES
Romania - Institutul National De Sanatate Publica	Romania	NO
Polish Platform of Medical Research	Poland	NO
BBMRI.pl – biobanks	Poland	NO
France Cohortes	France	YES
Hungarian eHealth Service Space (EESZT)	Hungary	NO

Table A1. List of contacted data hubs.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Annex II – Complete survey

HealthyCloud survey: Incentives and metrics for data deposition and data usage

- By ticking the following box you acknowledge that you have read the information on the Welcome page, you accept the conditions of this survey and that you are above 18 years old.
- Data hub name
- Role

1. Incentive systems for data sharing

- Do you use any incentives to promote the submission of data to your facility?* [Yes /No]
    - (If answered yes) Describe them [open text]
- Do you use any incentives to promote the reuse of the data hosted in your facility?* [Yes /No]
    - (If answered yes) Describe them
- Do you use any other type of incentives?* [Yes /No]
    - (If answered yes) Describe them
- Do you have knowledge about any other incentives that can be useful for data submission or reuse that you might not be using?* [Yes /No]
    - (If answered yes) Describe them

2. Metrics and measures of data deposition and data usage

- How do you measure data usage in your data infrastructure?* [multiple choices]
    - if you selected others, please specify. [open text]
- How do you measure data deposition in your data infrastructure?* [multiple choices]
    - if you selected others, please specify. [open text]
- How do you audit data requesters' actions?* [multiple choices]
    - if you selected others, please specify. [open text]
- How do you audit data access authorizations?* [multiple choices]
    - if you selected others, please specify. [open text]
- How do you make public the metrics regarding data deposition and usage?* [multiple choices]
    - if you selected others, please specify. [open text]
- Other comments [open text]

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

Annex III - Incentives free text answers

#	Incentives to promote the submission of data
1	Return results as part of the access policy.
2	- Increase biobanks visibility, collections and services displayed in the directory. - Submitting metadata increases the findability of their data/samples. - Converting data to OMOP or FHIR to join the federated system also increases visibility and makes their data more interoperable. - Reimbursement model for biobanks participating in the cohort.
3	Dissemination of shared data: use mobilized data for promotional purposes, providing outreach opportunities for data submitters.
4	Research funders to reuse the network as a requirement to receive funds.
5	Introducing regulation, guidelines and standards.
6	- Data hosting during the course of research projects. - Data sustainability upon project end. - Training and support for reproducible computational analyses. - Data quality and curation.
7	- Discoverable and available for reuse. - Recognition from the research community.
8	
9	Scientific collaborations.
10	
11	
12	- Novel research questions being answered. - Avoiding having to do the required pseudonymization and other necessary data handling procedures (the hub does it on behalf of the data provider).
13	
14	Give assistance to data sharing which is mandatory for projects funded by national and european calls.
15	
16	Promise feedback on what science can tell about people's genes.
17	Maintenance and legal ground for future uses.
18	Quality improvement reports to general practices.
19	
20	
21	Knowledge and expertise of partners, consultancy, existing governance arrangements.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

| 22 | In some research and infrastructure projects, sharing the data via XNAT.bmia.nl is included explicitly as one of the deliverables. |

Table A2. Answers from the question about data submission incentives.

#	Incentives to promote the reuse of the data
1	- Make UK biobank resources widely accessible. - Reduced access fees for low and middle income countries. - Cover costs of compute and data storage to support research aims.
2	Reduction of the numbers of steps that users must go through in order to request access to the data.
3	
4	Use the BioMedIT package free of charge if allow third party usage
5	Promoting national health databases.
6	
7	
8	
9	Foster scientific collaborations.
10	
11	
12	
13	
14	Valorization of the data and the skills of the teams that produced them through a common portal.
15	
16	To reuse the data is much quicker and easier than to gather new data. Making another or too many surveys with genedonors is bothering them. So new data collection is a commodity on its own and this must be used wisely.
17	
18	
19	Highlighting as best practice e.g. newsletters.
20	
21	Data can either be project based or can be used by others with guardians permission or can be core and accessed by all with existing governance.
22	

Table A3. Answers from the question about data reuse incentives.

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

HEALTHYCLOUD
Health Research & Innovation Cloud

#	Other type of incentives
1	
2	
3	- Providing data stewardship as a service - Crediting data stewards for supporting submissions hence promoting submission of quality datasets
4	"Make your data FAIR" program, providing extra funding to research project teams for making their data FAIR and accessible for reuse
5	
6	Data shared for scientific research is hosted free of charge.
7	Metadata of datasets is harvested to other services for increased discoverability, for example to OpenAire Explorer.
8	
9	
10	
11	
12	Ask researchers to participate in piloting of novel features or tools and may forgo some billing or speed up their application processing if necessary.
13	
14	
15	
16	
17	
18	
19	Has been used for project partner inclusion
20	
21	
22	I have published "data release papers" , describing datasets on XNAT.bmia.nl, which can be cited just like normal scientific publications.

Table A4. Answers from the question about other types of incentives.

#	Other incentives that can be useful for data submission or reuse
1	Highlight exemplar data returns and/or use of the resource, which can be helpful to researchers' career development/visibility.
2	

D4.3 Report on the usage and auditing of the data hubs indicators
and proposal of an incentive system
Version 2.0

3	Fund research projects that reuse data.
4	Requirements by publishers / journals.
5	
6	
7	
8	• Discounting access to services (or access to other data) in return for users who bring their data to the hub and make it available to others. • Also funders may insist that all data created through their research studies is made available through a hub.
9	
10	
11	
12	Perhaps exclusivity of use for some partners for some time before opening up the novel data to all researchers.
13	
14	Making data sharing a legal obligation, as the Health Data Hub in France was able to do, with mixed results, however.
15	• It would be a great incentive to make one's own data available if one could then also use other databases and possibly merge them with one's own data - especially in the scientific field. • The use of a uniform pseudonym or the provision of technical resources for generating a uniform pseudonym, as well as professional support and advice for data holders, are therefore very good incentives.
16	Advertising what good has come out of submitting data.
17	
18	Financial reimbursement.
19	Accreditation, highlighting as best practices, inclusion in projects.
20	Showcasing data to improve reuse and thus citation of the publications
21	Financial (e.g., cost-neutral hosting), negotiated requirement for sharing from funders
22	• Making the data better findable, by including metadata of the data on XNAT.bmia.nl in catalogues. We actually have done this for selected projects, but we haven't exploited and promoted this enough yet to really turn it into an incentive. • Long-term archiving for a fixed/low price, such that XNAT.bmia.nl is recognised as a true repository with persistent identifiers etc similar to figshare/zenodo.

Table A5. Answers from the question about other incentives that can be useful.