Baltic Data Flows

Co-financed by the Connecting Europe Facility of the European Union

MILESTONE 5 Progress Report

KPIs available for assessing the outreach of data

16 December 2021

This milestone progress report provides an overview of work achieved under Activity 2.3., 'Development and implementation of Key Performance Indicators (KPIs) to monitor the access and use of datasets'. KPIs are now active and will continue to monitor project impact until project completion in September 2023.

As stated in the Action Grant Agreement, *the task will monitor and assess the outreach of public open regionally harmonised datasets by developing KPIs to assess the re-use of datasets created under the Action and other data available from the analysed platforms. The KPIs, which will be specified in detail at the beginning of the Action, will answer to the questions:*

- *Who are the users of HELCOM Map and Data service?*
- *Which data themes of HELCOM Map and Data service are most used?*
- *What is the relative amount of traffic in HELCOM services originated from the EDP?*

Users will be classified in different groups. The KPIs will be presented in a user-friendly dashboard.

KPI development has been informed by findings from the HELCOM Map and Data Service (MADS) end-user survey conducted in March-April 2021, and the Spatineo Monitor, an online site monitoring tool that provides real-time usage analytics of the MADS. These two data collection tools are working together to monitor the project against eleven KPIs.

Analysis of results from 158 survey respondents of the end-user survey has been used to generate the baseline for KPIs 1-7. These will be measured against findings following a repeat of the end-user survey along with the project impact assessment survey (Activity 7.4) at the end of the project.

The Spatineo Monitor usage analytics tracks monthly unique visitors, data requests, and data transfers for each service available in the HELOCM MADS. Every month data from the Spatineo Monitor is used to assess monthly changes in user traffic and obtain a trends analysis on user engagement with the datasets available in the MADS. The Spatineo Monitor data is used for KPIs 8-11.

An additional KPI has been developed to monitor the number of datasets downloaded from the HELCOM Metadata Catalogue (KPI 12). This data will be made available from a customized download utility developed under Activity 2 of the BDF project during last quarter of 2021. KPI 12 data will be available for analysis in January 2022.

The KPIs are presented in this document and data is summarized in this Power-Bi dashboard. The dashboard currently presents KPIs from Quarter 2 2021 as the baseline (page 1) and will later present an update following the repeat of the End-user survey in Quarter 2 2023. The dashboard also presents the monthly usage analysis (unique visits, and requests) for each service (pages 2, 3, and 4), data transfer rates (page 5), and the number of referred data requests by origin (page 6), all collected via the Spatineo Monitor. Users of the dashboard can query data fields of interest by selecting the available visuals and filters.

Project Key Performance Indicators:

End-user survey baseline

1. Diversity of user groups
2. Utilisation of data by user location
3. Utilisation of data by service theme
4. Frequency of use
5. Purpose of data
6. Importance of data
7. Accessibility of data

Spatineo Monitor tracking

8. Number of data requests
9. Number of unique visitors
10. Amount of data transferred
11. Number of referred data requests

Metadata Catalogue reports

12. Number of datasets downloaded

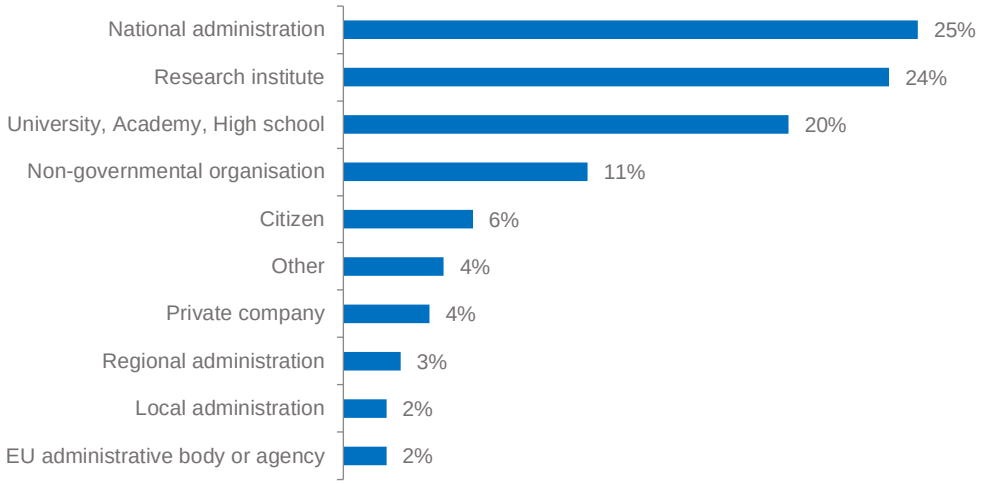
End-user survey baseline

1. Diversity of user groups

This KPI has been developed to specifically answer the question presented under Activity 2.3 of the project Grant Agreement; *Who are the users of HELCOM Map and Data service?* These groups have been initially identified through the end-user survey, with the baselines presented in figure 1.

Figure 1 visualizes the response data from question 1 of the end-user survey requesting that users indicate their professional background. Almost half of respondents represented only two professional backgrounds: 'National administrations' (25%), and 'Research institutes' (24%). There appears to be less engagement from the private sector (4%), regional and local administrations (5%), and EU administrative bodies. These groups will be assessed again with the repeat of the end-user survey and project impact assessment to see if HELCOM services are reaching a higher or lower diversity of user groups.

Figure 1: User groups

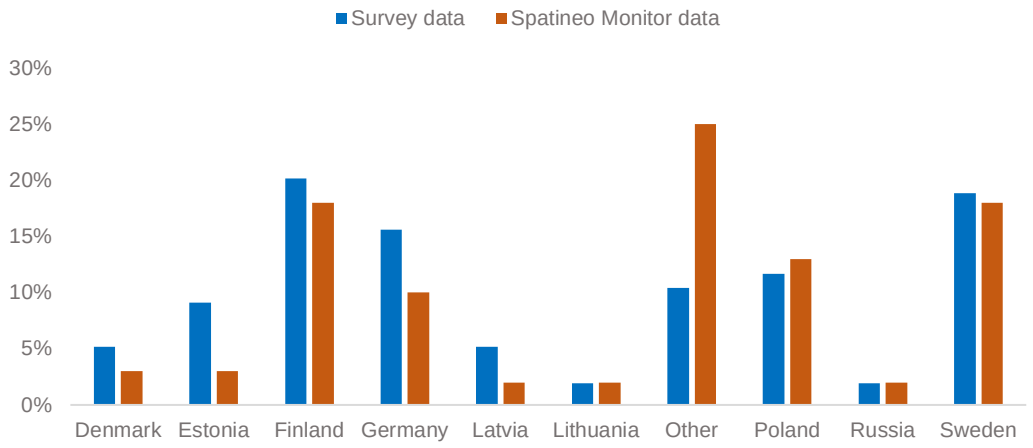


One of the principle aims of the project is to share datasets to a wider community of users such as European open data ecosystem, researchers, NGOs and the private sector. The benefits from the establishment of data harmonization will make data more accessible to these groups. It is therefore expected that a more diverse base of user groups will use these services as a result of the project.

2. Utilisation of data by user location

User location has been identified through the end-user survey in addition to analysis of log files from the Spatineo Monitor, as presented in figure 2. Figure 2 visualizes the response data from question 2 of the end-user survey requesting that users indicate which county they are based. Figure 2 shows that the highest number of respondents were from **Finland (20%), Sweden (19%), and Germany (16%)**, however several other countries outside of HELCOM are also benefiting from the MADS (10%). These findings also partly correspond with the data from the **Spatineo Monitor**, also presented in figure 2, for the reporting month of April 2021, the same month as survey data collection, with the top 3 countries of Finland (18%), Sweden (18%) and Poland (13%). However, the Spatineo Monitor reports a higher number of users in countries outside of HELCOM for the month of April (Other, 25%). These countries include the USA, UK, and France with over 100 users. The 'top 3' other countries reported in the survey include Netherlands, France, and Spain. The Spatineo Monitor indicates that the number of unique visitors to the HELCOM MADS during the month of **April 2021 was 4,624** (see KPI 8 for more details).

Figure 2: Users per country



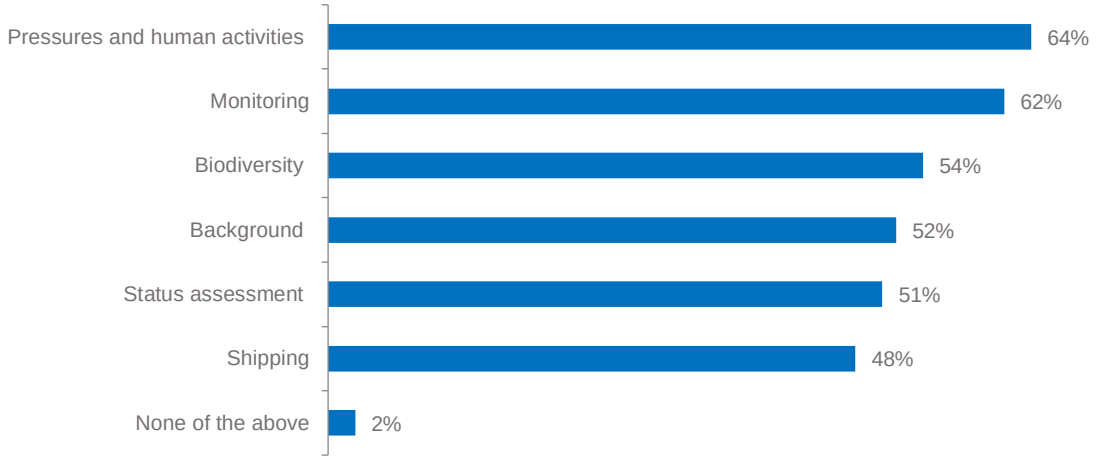
User location will again be collected with the repeat of the end-user survey and the project impact assessment survey. The expectation is that a broader number of users across the Baltic states would benefit from data services. The Spatineo Monitor log files will also again be used to provide an overview on the number of users and locations so a comparison between Quarter 2 2021 baseline can be considered. This IP address data will be anonymized, and no personal data will be collected using Spatineo's anonymization tool. Anonymization is achieved by removing information from the end of collected IP addresses, a process known as masking and ensures that usage analytics of services meet GDPR requirements.

3. Utilisation of data by service theme

Using this KPI data we can answer the question presented under Activity 2.3 of the project Grant Agreement; *Which data themes of HELCOM Map and Data service are most used?* This has been initially assessed through the end-user survey.

When the end-user survey was conducted in March-April 2021 data layers in the MADS were grouped into 6 services as shown in the 'Y' axis of figure 3. Figure 3 visualizes the response data from question 10 requesting that users indicate one or more service/s that they use. The most popular service selected was **'Pressures and human activities' (64%)**. This service includes distribution of different pressure/human activities effecting the marine environment (example dataset). This is closely followed by **'Monitoring' (62%)** that includes information on monitoring station networks and boundaries of HELCOM assessment unit areas (example dataset). MADS services are equally used across users with 94% of respondents indicating they use **2 or more services**, and 10% of respondents stating they use **all services**.

Figure 3: Data service/s used



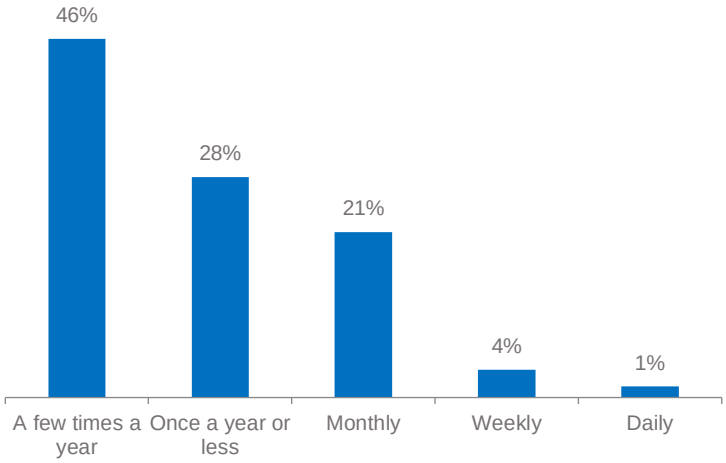
The Spatineo Monitor was also used to complement these findings by recording the number of visitors per MADS service by collecting the unique IP address of each user. Data for the month of April 2021, the same month as survey data collection, does not run exactly parallel to the MADS survey findings. The Spatineo Monitor highlights that '**Shipping**' is the most popular service with 810 unique visitors, followed by **'Background'** (785 users) and '**Background**' (784 users). However, as data from the survey shows, these services are used almost equally across users. Further presentation and analysis of the Spatineo Monitor unique visitor data is provided under KPI 9.

The utilization of services will again be assessed following the repeat of the end-user survey and project impact assessment in Quarter 2 2023. The expectation is that this will largely remain the same with the equal use of data services. It should be noted that the repeat of the survey will accommodate the restructuring of some MADS services in August 2021. The restructuring aims to improve dataset navigation and accessibility as the MADS continues to host more datasets. One additional service was created (Red listed species and habitats), the service 'Pressures and human activities' was split into two, and the service 'Status assessment' was renamed to 'Indicators and assessments'. There are now a total of 8 services.

4. Frequency of use

Based on the end-user survey findings, almost three quarters of respondents (74%) visited the MADS '**a few times a year' (46%)** or '**once a year or less' (28%)**. The Metadata Catalog was visited less often with respondents indicating 'once a year or less (58%), or 'a few times a year' (30%). The indicator baseline is presented in figure 5. Figure 5 visualizes the response data from question 3 of the end-user survey requesting that users indicate how often they visit the HELCOM MADS.

Figure 5: Frequency of use



The expectation towards the end of the project is that this will remain largely the same but with a shift towards a higher frequency of use as the project enhances the performance and functionality of the HELCOM MADS. Following the improvements throughout the project lifespan, along with the continued availability of new datasets, the expectation is for a higher proportion of monthly, weekly, and daily users.

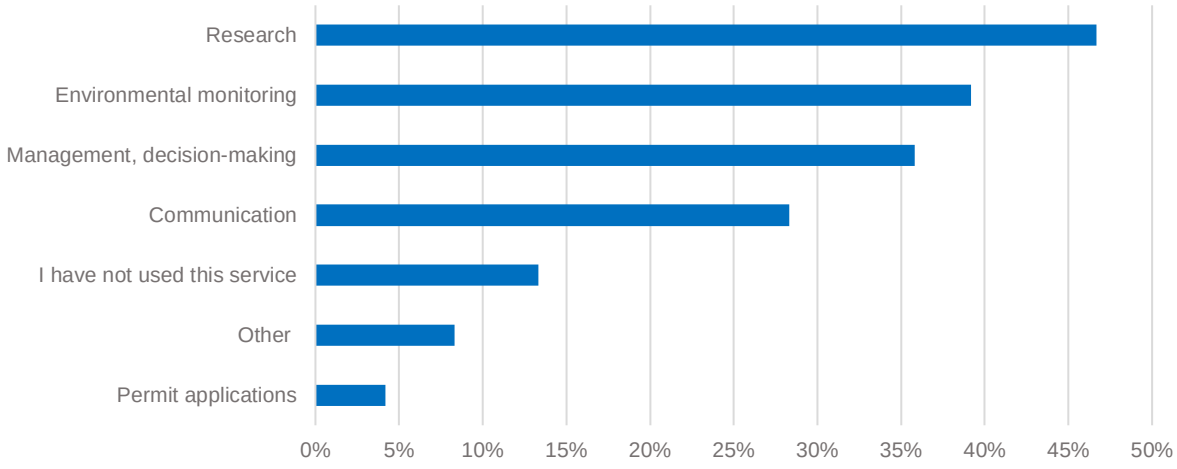
This KPI is closely linked with KPIs 8 (number of requests) and 9 (number of users) that monitor real-time data requests and unique visitor numbers rather than the user perception on frequency of service use. The Spatineo Monitor started monitoring MADS user traffic in December 2020. From January 2020 to April 2021 there was an **average of 5,150 users per month** to the MADS. The average number of users per quarter is further presented and analyzed at the service level in KPIs 8 and 9 detailed in this report.

5. Purpose of data

Figure 6 visualizes the response data from question 5 of the end-user survey requesting that users indicate for what purpose they use the HELCOM MADS. Most respondents use the MADS for **'Research' (47%), 'Environmental monitoring' (39%),** and **'Management, decision-making' (36%)**. The repeat of the end-user survey along with the project impact assessment will consider this baseline to further understand if the data purpose has been extended across multiple cases, or if it has narrowed in purpose following outputs of the project.

The expectation is that data will be used for a broader variety of purposes as the HELCOM MADS continues to expand and benefit from data harvesting. In addition, 'Communication' purposes will become more important as the MADS engages with a wider community of users through dissemination of data products and tools.

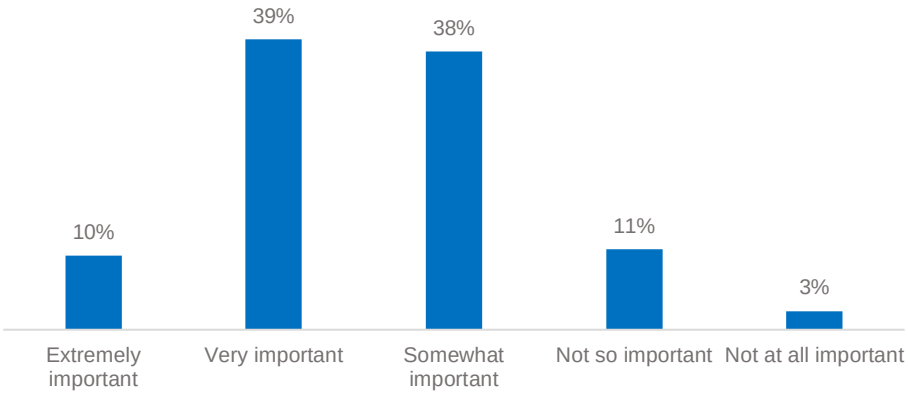
Figure 6: Purpose of data



6. Importance of data

Figure 7 visualizes the response data from question 6 of the end-user survey requesting that users evaluate the importance of HELOCOM MADS. Figure 7 highlights that almost half of respondents (49%) consider the MADS as either **'Extremely important' (10%) or 'Very important' (39%) in terms of content and importance of the data accessed from MADS**. Using this as a baseline the project impact assessment will highlight if data services have increased or decreased in status and importance.

Figure 7: Importance of data

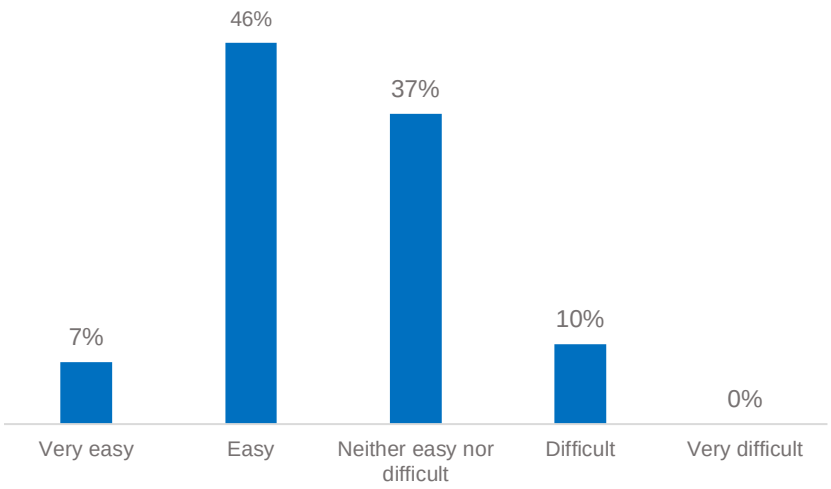


The expectation towards the end of the project is this will remain largely the same but with a slight shift towards higher importance as more users rely on the data provided by MADS. This will likely be achieved through the automation of data harvesting systems based on Application Programming Interfaces (APIs) developed with the aim to automatically integrate national datasets into a combined and harmonised regional dataset.

7. Accessibility of data

Figure 8 visualizes the response data from question 13 of the end-user survey requesting that users indicate how easy it is to find data on HELOCOM MADS. Over half of respondents (53%) indicated that it was **'Very easy' (7%) or 'Easy' (46%)** to find data from the MADS. The repeat of the end-user survey along with the project impact assessment will identify if the project has facilitated access to data services or not during the MADS update phase carried out under Activity 2 during 2021-2022. The expectation is that data will become easier to access with the upgrade of both the MADS and the linked HELCOM Metadata Catalog.

Figure 8: Accessibility of data



Spatineo Monitor tracking

KPIs 8-11 do not draw on respondent data from the end-user survey. Instead, these KPIs are reliant on monitoring of MADS services using the Spatineo Monitor. The usage analytic data for each of the eight services is presented on pages 2-6 of the KPI Power-Bi dashboard. The dashboard presents usage analytics for both Arc GIS ESRI server and WMS service types however KPIs will only consider usage data of the Arc GIS ESRI server, the origin of all MADS user requests.

In addition, as noted in KPI 3, MADS services were restructured in August 2021. One additional service was created (Red listed species and habitats), the service 'Pressures and human activities' was split into two, and the service 'Status assessment' was renamed to 'Indicators and assessments'. There are now a total of 8 services. As a result of restructuring the Spatineo Monitor started following these services in September 2021. This should be considered when using the data and graphs presented in KPIs 8-11.

8. Number of data requests

Each service available in the MADS has different levels of use and popularity every month. The Spatineo Monitor makes it possible to analyze these trends over time. Using this data we can again answer the question presented under Activity 2.3 of the project Grant Agreement; *Which data themes of HELCOM Map and Data service are most used?*

Figure 9 highlights the trend in the number of data requests per month for each service available on the HELCOM MADS. The trend in data requests over the year 2021 is roughly the same for each service with a peak in data requests in March, followed by a gradual decline in requests towards June, July, and August, and then with an upward trend in September and October 2021. The peak experienced in March is likely linked to an increase in user traffic with the launch of the end-user survey, and the down period in the summer months a result of the European summer vacations. The increased number of requests in September and October might be an indication of increased popularity of the MADS in recent months, however this will not be clear until all data is received for quarter 4 2021.

Figures 9 and 10 highlight that **'Biodiversity' received the highest amount of data requests at 357,615 requests for 2021 so far.** Figure 9 shows 'Biodiversity' reached a **maximum of 60,730 in March, and a minimum of 12,136 in August**. As shown in figure 10, 'Background', 'Pressures' and 'Shipping' have **all received around 270-280K requests for 2021 so far**. The restucturing of MADS introduced the bottom 3 services shown in figures 9 and 10 when the Spatineo Monitor started to provide usage analytics in September 2021. It is therefore not possible to make a compasrsion for the total number of data requests until we have collected annual data for all services. This will be possible in 2022 and 2023. For a more detailled summary of the data, view the Power Bi dashboard that seperates each service and overlays requests and unique visitor numbers (KPI 9) within individual graphs (pages 2 and 3).

Figure 9: Number of data requests for each MADS service in 2021

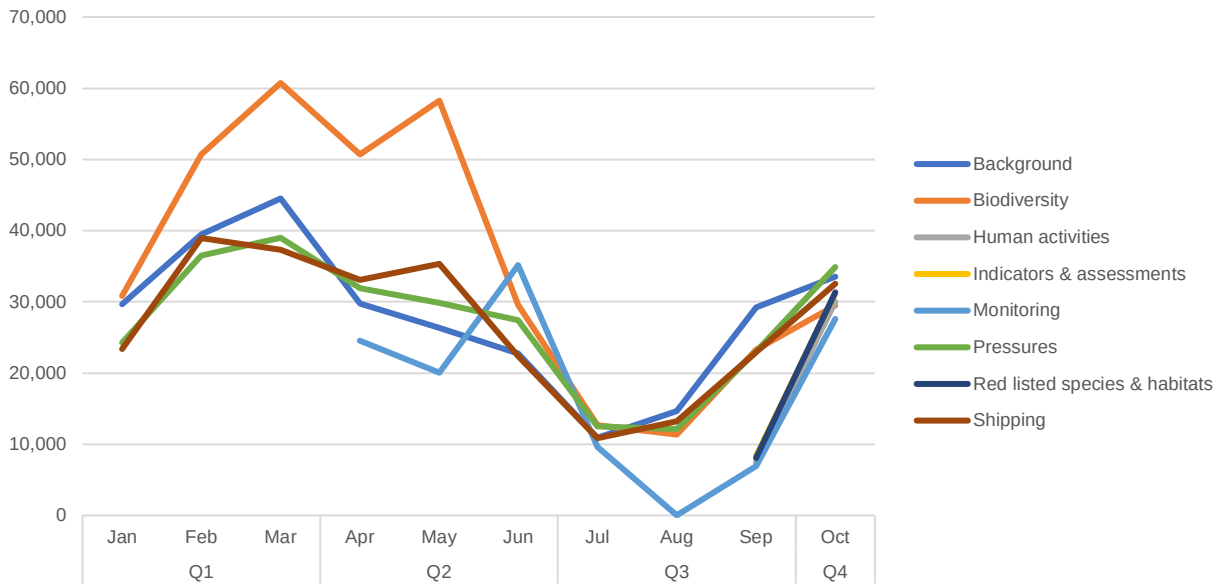
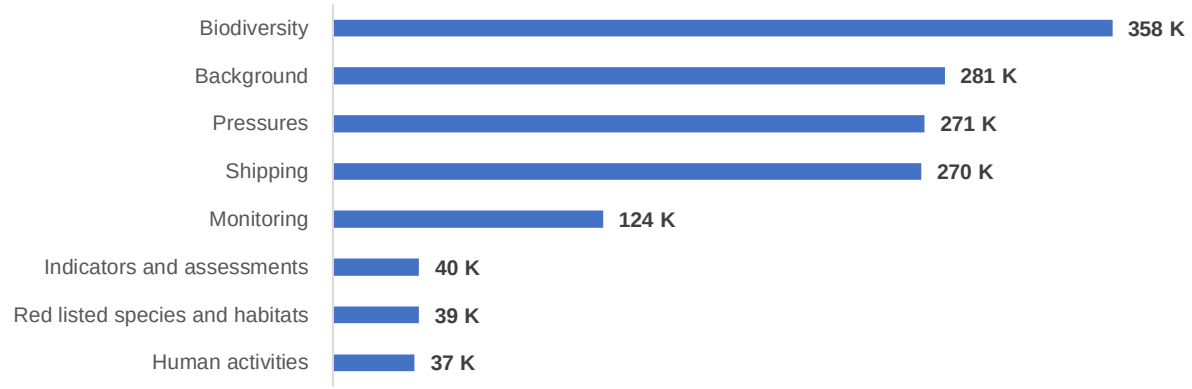


Figure 10: Total number of data requests per service in 2021



In addition, the Power BI dashboard presents KPI visuals showing the number of data requests on each service per month. It takes the rolling yearly data request average since January 2021 for each service and highlights if the most recent month has received a higher or lower amount of data requests in comparison to that average. As shown in figure 11, **all services in October 2021, except 'Biodiversity', received an above average number of data requests**, indicated using the green colour and tick icon. The **'Biodiversity' service has the highest yearly average at 36,458 data requests**. The trend in data requests can also be seen in the shape of KPI visual background, in either green or red colour. The month can be changed using the 'month' filter available in the dashboard.

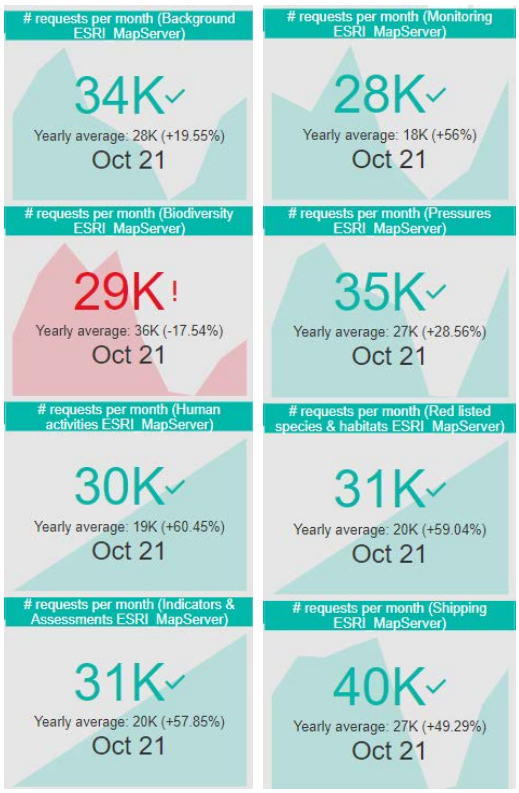


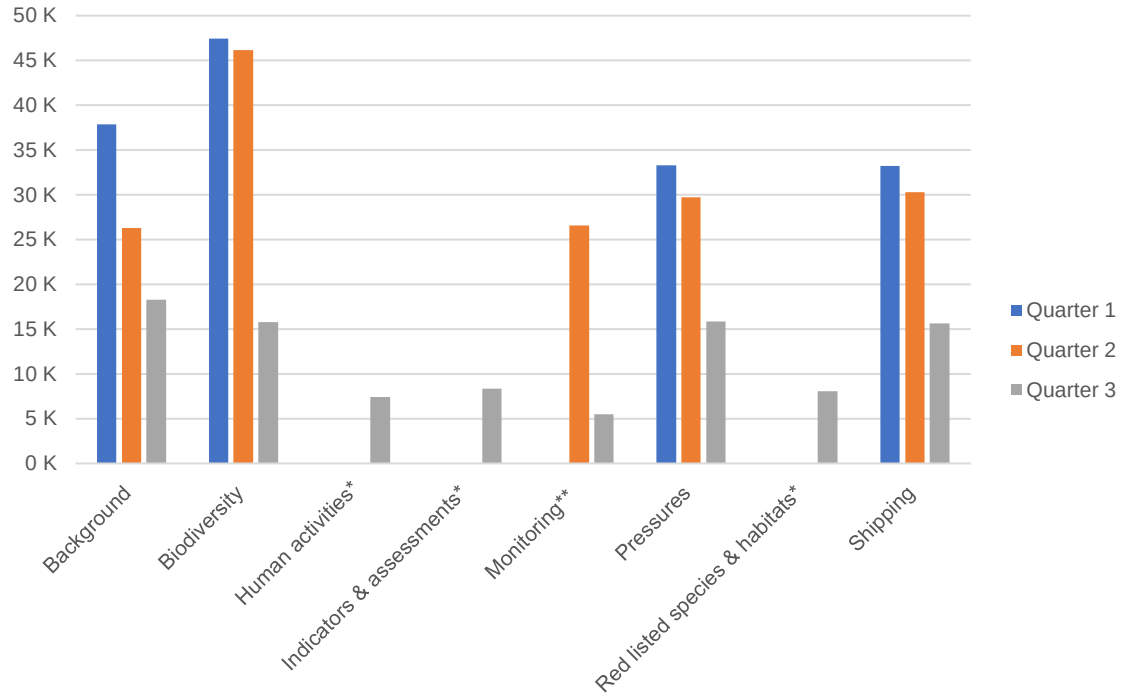
Figure 11: snapshot from Power BI dashboard showing KPI visuals on the number of data requests per service in October

To effectively assess the impact of the project over a relevant timeframe, KPIs will also consider the average number of data requests every quarter for each service, as highlighted in table 1, figure 12 (currently only for quarters 1-3), and on page 4 of the Power BI dashboard. This will reduce the effects of daily, weekly, and monthly request fluctuations due to seasonal holidays and is also a more effective grouping for trend analysis over the project duration. The project impact assessment under activity 7.4 at the end of the project will consider these quarterly averages to evaluate if the number of data requests have increased or decreased over the project duration. The project is scheduled to finish in September 2023 and so yearly averages will also be available for the impact assessment. The same quarter for each year of the project (2021-2023) will be compared and the expectation is that the MADS will receive more data requests for each year.

Table 1: Average quarterly data requests for each service in 2021

Service	Q1	Q2	Q3
Background	37,879	26,267	18,255
Biodiversity	47,405	46,185	15,786
Human activities*	-	-	7,400
Indicators & assessments*	-	-	8,332
Monitoring**	-	26,586	5,528
Pressures	33,263	29,718	15,882
Red listed species & habitats*	-	-	8,068
Shipping	33,229	30,262	15,652

Figure 12: Average number of data requests per quarter for each service in 2021



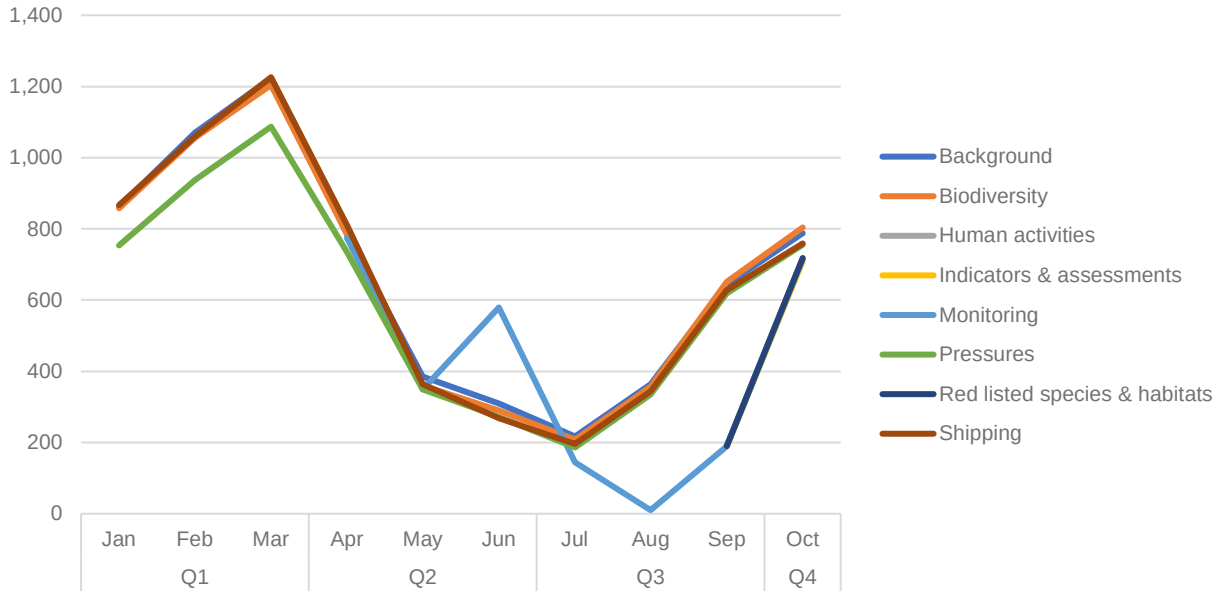
Following the restructuring of the HELCOM MADS services in August 2021, services marked with (*) are only monitored from September 2021 onwards, and those marked with (**) from April

9. Number of unique visitors

The Spatineo Monitor captures the number of unique visitors by collecting the IP addresses of MADS users. The IP address data is anonymized, and no personal data is collected. IP address anonymization is achieved by removing information from the end of collected IP addresses, a process known as masking and ensures that usage analytics of services meet GDPR requirements. This is a useful monitoring tool and informs KPI 9 on a monthly, quarterly, and yearly basis. Automated reports can be generated in the Spatineo Monitor to provide this data within a specific timeframe.

Figure 13 highlights the trend in the number of unique visitors per month for each service available on the HELCOM MADS combined into one graph. As reflected in the number of data requests presented in KPI 8, the number of unique visitors follows the same trend in 2021 with a spike in usage coinciding with the launch of the end-user survey, a slowdown in the summer months, followed by an upward trend starting from September. However, the peak in March is different to KPI 8 with maximum number of unique visitors to the **'Background' service (1,223 visitors) instead of 'Biodiversity'**. 'Background' service has received 6,648 visitors in 2021 so far, followed by 'Biodiversity' (6,570), and 'Shipping' (6,518).

Figure 13: Number of unique visitors per month for each service



The Power BI dashboard also presents KPI visuals showing the number of unique visitors on each service per month. Similar to KPI 8 on number of data requests, it takes the rolling yearly average since January 2021 for each service and highlights if the most recent month has received a higher or lower number of unique visitors in comparison to that average. As shown in figure 14, **all services in October 2021 received an above average number of unique visitors**, indicated using the green colour and tick icon. The **'Background' service has the highest yearly average at 665 unique visitors**. The trend in data requests can also be seen in the shape of KPI visual background, in either green or red colour. The month can be changed using the 'month' filter available in the dashboard.

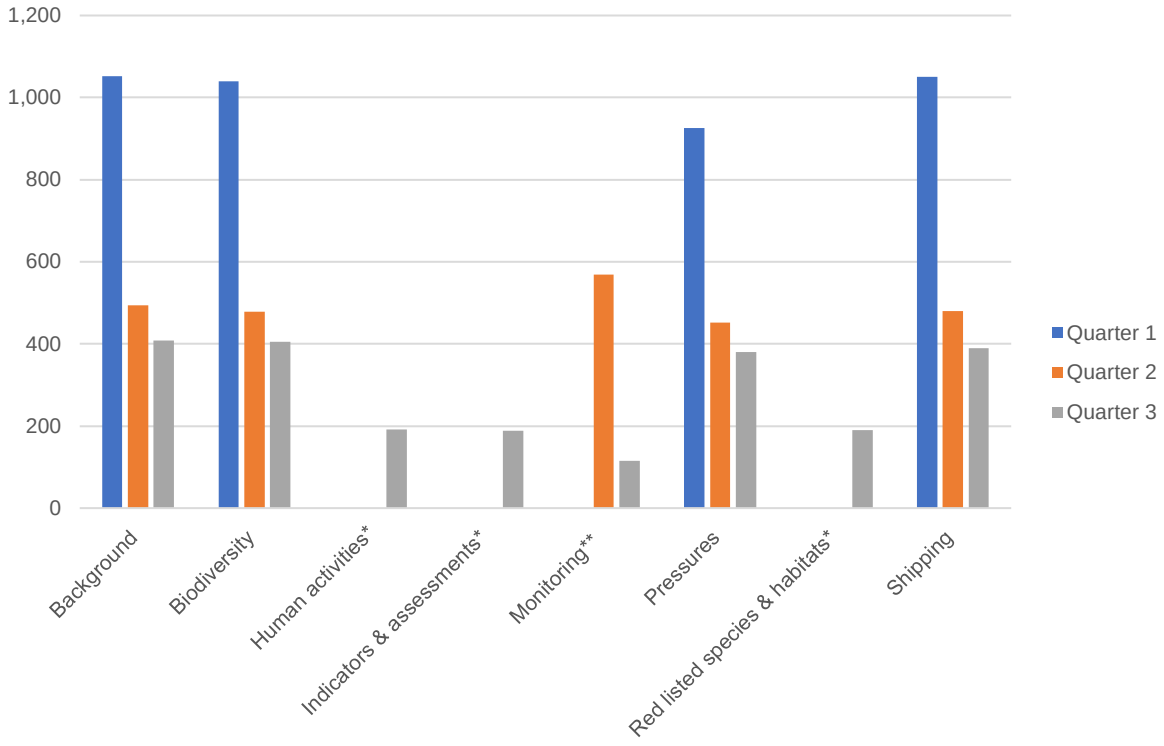
Figure 14: snapshot of Power BI dashboard showing KPI visuals on number of unique visitors per service in October

Similar to KPI 8, KPI 9 will also consider the average number of visitors every quarter for each service, as highlighted in table 2, figure 15 (currently only for quarters 1-3) and on page 4 of the Power BI dashboard. For those services monitored since January 2021, the **highest average number of users was significantly higher in quarter 1, compared to quarters 2 and 3**. The project impact assessment under activity 7.4 at the end of the project will consider these quarterly averages to evaluate if the number of visitors has increased or decreased over the project period. The project is scheduled to finish in September 2023 and so yearly averages will also be available for the impact assessment. The same quarter for each year of the project (2021-2023) will be compared and the expectation is that the MADS will receive an increasing and upward trend of visitor numbers.

Table 2: Average number of users per quarter for each service in 2021			
Service	**Q1**	**Q2**	**Q3**
Background	1,052	493	408
Biodiversity	1,039	478	405
Human activities*	-	-	191
Indicators & assessments*	-	-	189
Monitoring**	-	568	115
Pressures	926	452	380
Red listed species & habitats*	-	-	190
Shipping	1,050	480	389

Following the restructuring of the HELCOM MADS services in August 2021, services marked with (*) are only monitored from September 2021 onwards, and those marked with (**) from April 2021

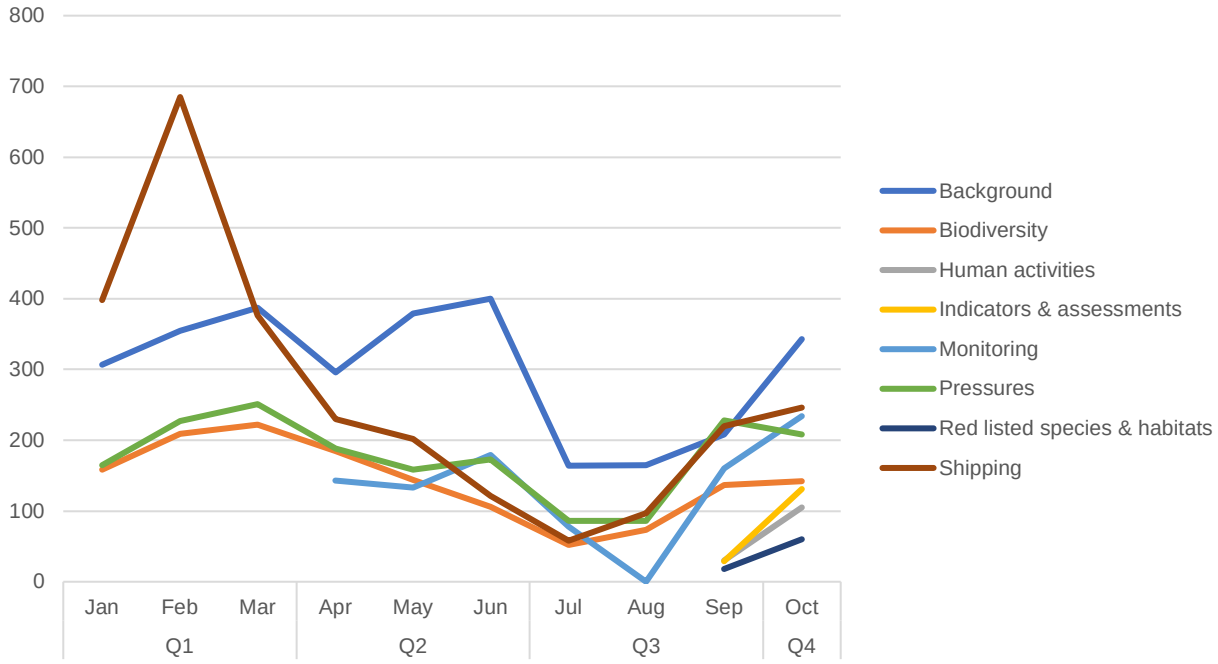
Figure 15: Average number of visitors per quarter for each service in 2021



10. Amount of data transferred

The Spatineo Monitor also monitors the amount of data transferred per service recorded as the number of mebibytes (MiB) transferred per month. Along with KPIs 8 and 9, this is another indication on the popularity of MADS services, but instead shows how much data is transferred or downloaded using the MADS. As shown in figure 16, the trend in data transfers is slightly different compared to the number of data requests and unique users and is more stable across the year. There was no peak coinciding with the launch of the MADS end-user survey indicating that users were simply viewing the service but not actually downloading data at the time. There is a decrease in the number of data transfers in the summer months and an upward trend starting from September. Similar to data presented in KPI 9, the 'Background' service is the most popular with the highest monthly transfer rate. However, there is one exception; 'Shipping' experiences a very high data transfer rate in February (685 MiB).

Figure 16: Montly data tranfers (MiB) for each service in 2021



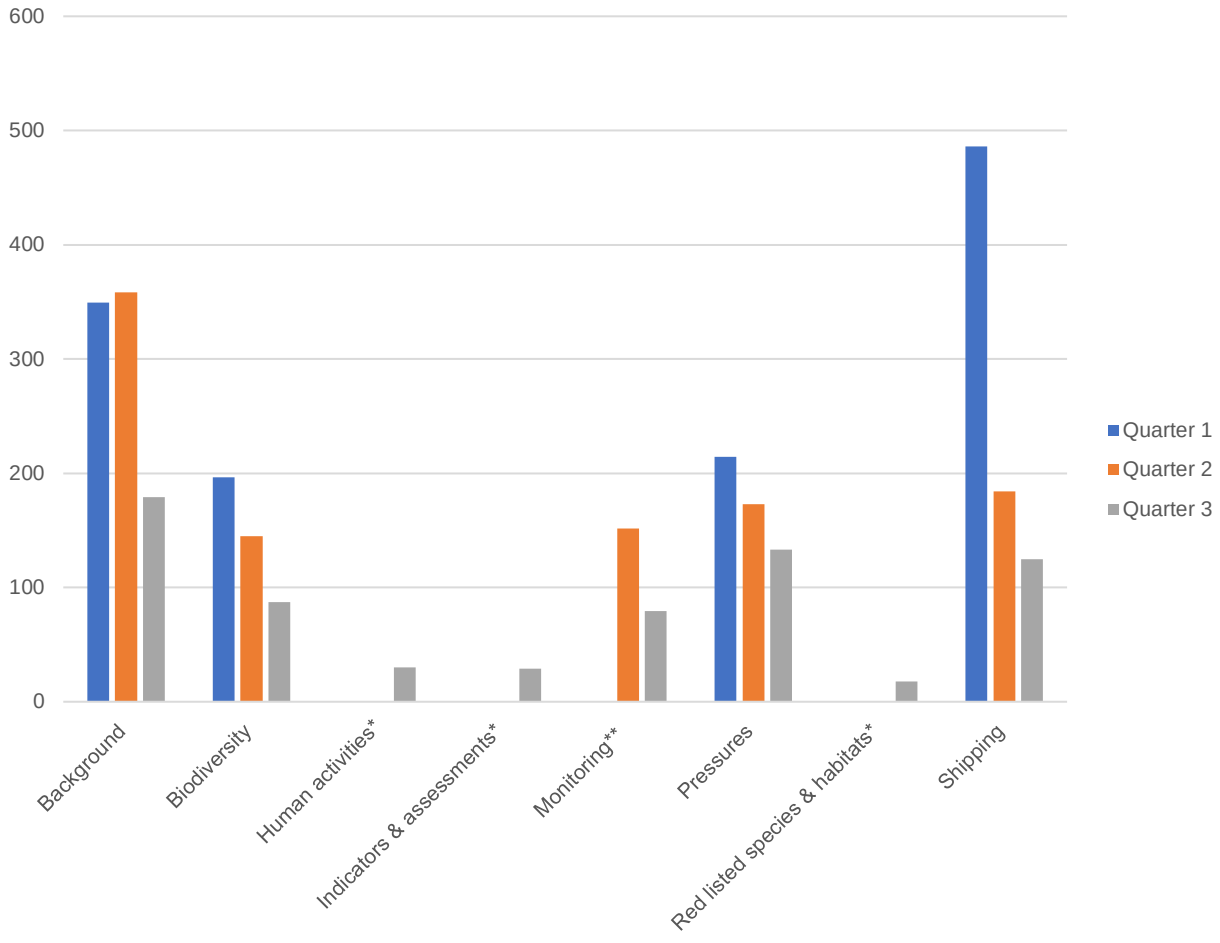
Similar to KPIs 8 and 9, KPI 10 will also consider the average data transfer rate every quarter for each service, as highlighted in table 3 and figure 17 (currently only for quarters 1-3). For those services monitored since January 2021, the highest average data transfer rate was experienced in quarter 1, except for 'Background'. For a more detailled summary of the data, view page 5 of the Power Bi dashboard.

The project impact assessment under activity 7.4 at the end of the project will consider these quarterly averages to evaluate if the transfer rate has increased or decreased over the project period. The project is scheduled to finish in September 2023 and so yearly averages will also be available for the impact assessment. The same quarter for each year of the project (2021-2023) will be compared and the expectation is that the transfer rate will increase indicating more users are accessing and downloading data from MADS.

Table 3: Average data transfers (MiB) per quarter for each service in 2021			
Service	**Q1**	**Q2**	**Q3**
Background	350	358	179
Biodiversity	196	145	87
Human activities*	-	-	30
Indicators & assessments*	-	-	29
Monitoring**	-	152	79
Pressures	214	173	133
Red listed species & habitats*	-	-	18
Shipping	486	184	125

Following the restructuring of the HELCOM MADS services in August 2021, services marked with (*) are only monitored from September 2021 onwards, and those marked with (**) from April 2021

Figure 17: Average data transfers (MiB) per quarter for each service in 2021



11. Number of referred data requests

Using the Spatineo Monitor we can track how data requests and visitors are referred to the HELCOM MADS services. This is important to gauge the effectiveness of implementing data harvesting platforms under the project to improve access to datasets. In particular, the project will make available environmental metadata records via the 'data.europa.eu' open data platform and this KPI has been designed to specifically answer the question presented under Activity 2.3 in the project Grant Agreement; *What is the relative amount of traffic in HELCOM services originated from the EDP?*

Table 4 shows a list of the top 10 referrers by total number of data requests for all HELCOM MADS services since 1 January until 30 October 2021. As recorded by the Spatineo Monitor, **the HELCOM MADS domain (maps.helcom.fi) is the main entry point to services and is listed as the top referrer for 96.8% of data requests**. In order to effectively monitor the impact and prevalance of other referrers the 'maps.helcom.fi' has been excluded from this analysis. However the percentage of **data requests that originate from referrers will be considered, currently at 3.2%**. Page 6 of the Power Bi dashboard provides an overview of all refferrs per service for 2021. One or more of the referrer domains can be selected from the filter 'Select service referrer'. The dashboard will later include data for 2022 and 2023 so a comparative annual analysis can be achieved.

Table 4: Top 10 referrers to MADS services, Jan-Nov 2021

Referrer	# data requests
portal.ox2.com	14,915
rhk.maps.arcgis.com	7,996
www.arcgis.com	4,949
Localhost (HELCOM Secretariat)	2,132
www.esri.com	2,038
storymaps.arcgis.com	1,729
geo.mil.intra	1,520
eibs.maps.arcgis.com	604
demo-kartta.paikkatietoikkuna.fi	579
helcom.maps.arcgis.com	518

The project impact assessment under activity 7.4 at the end of the project will consider this KPI to understand if the number of referrers, and data requests originating from referrers, is increasing, or decreasing. The expectation is that a broader diversity of referrers will appear as the project makes progress with implementation of data harvesting. In addition, more data requests will originate from referrers, increasing the percentage of 3.2% that are currently referred in comparison to access directly from 'maps.helcom.fi'. In addition, the 'data.europa.eu' platform will appear listed as a referrer starting in 2022.

12. Number of datasets downloaded

This data will be made available from the customized download utility of the HELCOM Metadata Catalogue. This utility has been developed under Activity 2.1 of the BDF project during last quarter of 2021. KPI 12 data will be available for analysis in January 2022.