UNITED
NATIONS

EP

UNEP/MED WG.566/9

UNEP

**UNITED NATIONS
ENVIRONMENT PROGRAMME
MEDITERRANEAN ACTION PLAN**

10 May 2023
Original: English

Meeting of INFO/RAC National Focal Points

Rome, Italy, 7-8 June 2023

Agenda item 5: MAP Data Policy - Data flows Annexes

Data policy Annex - InfoMAPNode data flow

In charge of the study at INFO/RAC

Lorenza Babbini, Director

Arthur Pasquale, Deputy Director

Alessandro Lotti, InfoMAPNode officer

Annalisa Minelli, Knowledge Management & Digital Transformation officer

Report prepared by:
Lorenza Babbini, Alessandro Lotti, Annalisa Minelli, Arthur Pasquale

Data policy Annex - InfoMAPNode data flow

Introduction

This document describes in details references and procedures tied to InfoMAPNode optional data flow with respect to the adopted Data Policy (UNEP/MED IG.25/27, Decision IG.25/10). It could be considered as a guideline to fulfill geographical data sharing necessities for Contracting Parties focusing on data flux structure, data types, access levels, data sources, data formats, data quality, data licenses, metadata, data sharing practices and restriction.

Moreover, InfoMAPNode will be the geographical core, identified as the "Data Hub", of the upcoming Knowledge Management Platform. This adds value to this flux, though not mandatory.

Summary of data management aspects

Brief description of the structure

Data is managed by INFO/RAC in the frame of a tripartite structure that is composed by a **Geoserver instance**, where only INFO/RAC members have access, a **Geonode instance**, where all data publishers have access in read and write mode for what concerns their own data, a **Pycsw-based structure** useful to harvest data from different online sources and keep data up-to-date, where only INFO/RAC members have access.

The architecture of InfoMAPNode is shown in Figure 1.

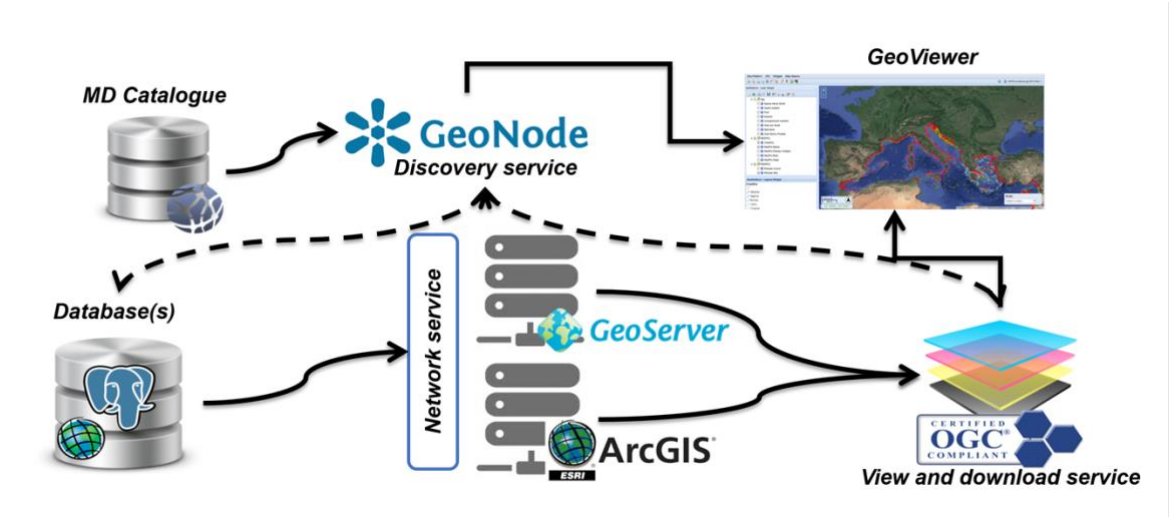

Figure 1 - Architecture of InfoMAPNode

A brief description of the individual components of the architectural scheme follows:

Geoserver: It is an open source server for sharing geospatial data. Designed for interoperability, it publishes data from any major spatial data source using open standards. GeoServer supports a variety of data formats and protocols, including WMS, WFS, WCS, and GeoJSON.

Geonode: It is an open-source geospatial content management system for sharing and visualizing geospatial data, with a web-based interface for users to upload, organize, and share geospatial data and maps. Users can create interactive maps, download data, and collaborate with others on geospatial projects.

Pycsw-based structure: PyCSW (Python Catalog Service for the Web) is an open-source, OGC-compliant catalog server for publishing and managing geospatial metadata. It provides a web-based interface for querying and harvesting metadata from various sources, such as spatial databases, file systems, and other web services.

LDAP server: LDAP (Lightweight Directory Access Protocol) provides a central place for authentication, storing usernames and passwords and validating them with services. System administrators can also use LDAP single sign-on to control access to an LDAP database.

InfoMAPNode is designed with five levels of access (see following section on Access Levels). INFO/RAC members have access to all contents. Furthermore, specific groups and teams can be created. Each user can be linked to one or more groups.

The general workflow starts with data ingestion into the portal, proceeds with data stylization (where available depending both on data type and access level), then the metadata are compiled and the finally data is published.

After login into the platform, through the homepage, a user can start uploading datasets (geographic layers) and documents or creating a map. Furthermore, using "Remote services" it can be imported remote services (WMS, Rest GIS etc) from other SDIs that use OGC standards. The general workflow connected to inputs and the products offered is reported in Figure 2.

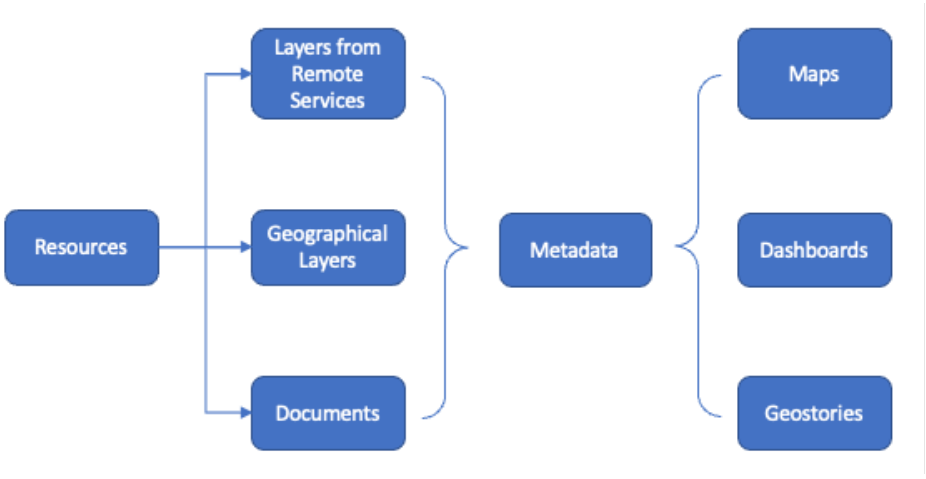

Figure 2 - Workflows and products in InfoMAPNode

Data management common practices among different access levels

InfoMAPNode relies on a virtual machine hosted on a physical server in ISPRA (Italy). The backup operations, with a double restore point, are executed daily. These practices are not foreseen to change in time, since they belong to the wider infrastructure (ISPRA) hosting INFO/RAC and the policy is reputed valid.

The actual software underlying the platform is Geonode 4.0 and Geoserver 2.20, recently upgraded to respond to the highest standards of security and the best availability of software (plugins) and assistance in time. The maintenance of the software is not yet fixed, but a yearly check is considered a good compromise between update necessities and man-hours requirements. Nevertheless, specific bugs and faults will be promptly amended. The personnel responsible for these actions is the INFO/RAC group plus IT personnel from ISPRA.

The group responsible for the platform is also responsible for data management and the executor of data management plan implementation. The UNEP-MAP data management plan and the document will be reviewed once per year in reason of new/upcoming modification on data sources, data structures, data sharing techniques and tools. to connect with the help desk of infoMAPNode, the following address is available: infomapnode@info-rac.org

I. Data Source characterization

InfoMAPNode currently canalizes data fluxes from different data sources. The principal two data fluxes are layers uploaded in the underlying instance of Geoserver and layers being fetched by a certain amount of remote services. At the present time, 98% of data comes from remote services and InfoMAPNode mirrors that data which is stored in Geoserver instances of SPA/RAC and GRID Geneva (managing data for Plan Bleu and Medwaves RACs). Only a small part of available data is stored in the Geoserver instance behind InfoMAPNode, following the **non-duplication of efforts and resources** principle. This implies that the availability of layers is strongly related to the persistence of the data source in particular, and the FAIR data management principles application by these sources in general.

In fact, the persistence of the identifier (of the remote service, in this specific case) is one of the pilllars of the Findability of a resource. The application of this specific principle grants the stability of InfoMAPNode infrastructure.

The data sources from UNEP-MAP already present on the InfoMAPNode platform are:

a. *REMPEC Medgismar dataset*: in this case, 4 layers, publicly available as .csv, have been imported as vector layers into the Geoserver instance underlying InfoMAPNode. The layers have been stilyzed and an .sld file has been attached to them. The layers are available for download in the same exact way as they are available in Medgismar viewer (https://medgismar.rempec.org/). For what concerns restricted access data, it has been decided in accord between REMPEC and INFO/RAC to don't import the layers, since they contain sensitive material (in terms both of privacy and legal issues connected with the information). The treatment of sensitive data will be matter of future decisions, also according to GDPR regulation and a literature recognition for what concerns the privacy-related matters in the whole Mediterranean area. Due to the fact that raw data has been often reworked by the RAC, data contained in this data source is owned by REMPEC and the Contracting Parties.
b. *SPA/RAC Geoserver instance*: reachable at http://tomcat.medchm.net/geoserver/ows?, the service counts for 297 layers from different projects leaded by SPA/RAC. The instance is fully compliant with Geonode and configured in a way that resources can be imported in any other system (included InfoMAPNode) as remote service using OGC standards. At the present time 240 layers from this Geoserver instance have been imported into InfoMAPNode, metadated and appropriate licenses (agreed between SPA/RAC and the Contracting Parties) have been transmitted to the mirrored layers. The layers contained in this instance are not downloadable at the present time. Due to the fact that raw data has been often reworked by the RAC, data contained in this data source is owned by SPA/RAC and the Contracting Parties.
c. *UNEP-GRID (Geneva) Geoserver instance, containing layers from Plan Bleu and Medwaves (before SCP/RAC):* reachable at https://geoserver.mapx.org/geoserver/ows?, the service counts for 168 items, 70 of which belong to UNEP-MAP components Plan Bleu (40 layers) and SCP/RAC (30 layers) respectively. Fully compliant with GeoNode, selected layers from this repository have been widely metadated by data owner and manager and mirrored in InfoMAPNode using OGC standards. The data manager has also relased the endpoints for each single layer of interest. Data within this instance does not have a license. The layers contained in this instance are not downloadable at the present time. Due to the fact that raw data has been often reworked by the RAC, data contained in this data source is owned by Plan Bleu, Medwaves and the Contracting Parties.

No data coming from the aforementioned sources is actually under **embargo**, nevertheless InfoMAPNode has the capabilities to deal with confidential data by appropriately setting sharing parameters according to necessities. It is a specific choice to not host, for the moment, data which is sensitive in terms of privacy or involved in legal issues.

As a general rule, data present in the sources described above is collected by contracting parties or stakeholders commissioned from contracting parties. Data is then provided from the CP to the RAC, that often harmonizes and sometimes reworks data to achieve a meaningful data representation, an adequate quality level or elaborate indexes and other derivative parameters of interest for UNEP-MAP. Subsequently, data is shared under RAC supervision. RACs are also charged for data quality assessment and control, and they are responsible for the quality of shared data.

The appointmentof InfoMAPNode Users is on voluntary basis, but it would constitute an asset for many aspects and will support the creation of always more diffused network at national level to facilitate data sharing and implementation of a spatial data infrastructure and shared environmental information system at Mediterranean level.

II. Licenses

According to the Data Policy, approved during the Convention for the Protection of the Marine Environment and the Coastal Region of the Mediterranean (Barcelona Convention) and its Protocols at their 22nd Meeting (UNEP/MED IG.25/27, Decision IG.25/10, p. 353), data must be as open as possible, respecting the constraints imposed by local legislation, sensitivity of data, and copyrights. For files published in or mirrored by InfoMAPNode by the RACs the principal Open Data license individuated by the policy is Creative Commons Attribution (CC-BY). Nevertheless, if data is originally shared by the data owner (which corresponds in most of the cases with data creator) with a less restrictive license (like CC-0 or Public Domain) or more restrictive licenses (such as CC-BY-NC or CC-BY-ND) it is necessary to transmit this license while data is shared following the workflow summarily described in Figure 3a (specific case of data owner coincident with the Country). Other, more restrictive, licenses are still available to use and correspond to specific access level of the users or sensitivity of data themselves. Figure 3b shows all the possible licenses from the open ones (bright green zone) to the classic copyright (in red), which denotes closed data. Given what is stated in the data policy, the usage of “out of bright green” licenses is regulated and it must be appropriately motivated in a participative process that involves data owner, eventual data provider (if any), and the subject responsible for data sharing (such as the RACs). For more references on which cases represent an exception to the general open data rule see the UNEP/MED IG.25/27, Decision IG.25/10 (Data Policy) at Section 4, Article 36.

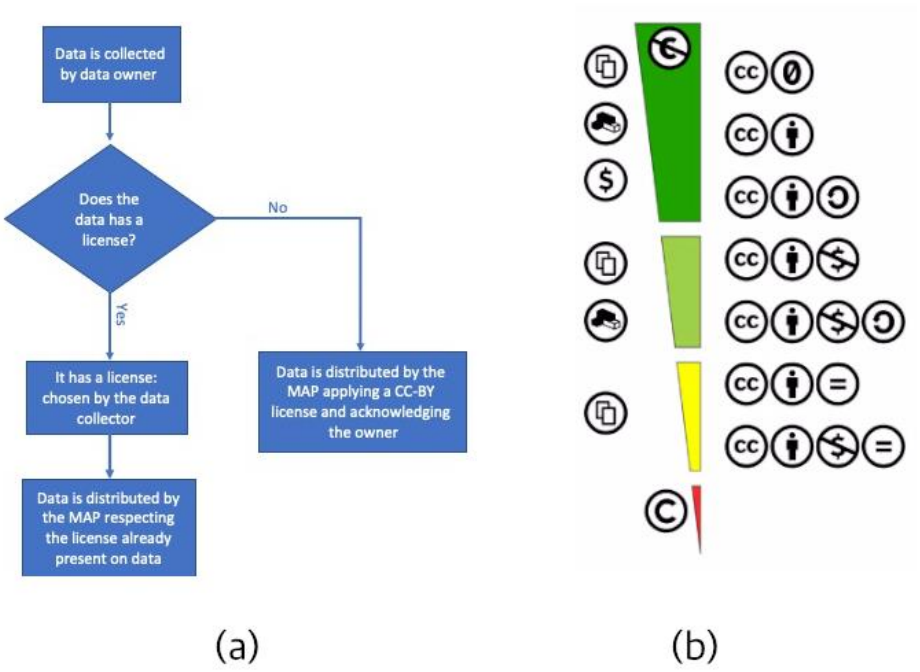

Figure 3 - License attribution fluxes for data (a); available licenses for data from bright green (open data) to red (closed data, b).

Production of metadata and other relevant documentation

Metadata are also an important part of data shared via InfoMAPNode. The metadata filling process starts as soon as a new resource (layer or remote service) is added to the platform. Metadata compilation could be pursued in InfoMAPNode both via a form to be filled in and a guided procedure evidencing required and optional parameters. For the moment, required parameters for an adequate level of metadatation (compliant with ISO 19155 standard) are:

- Dataset title
- Abstract (dataset description)
- Category
- Eventual group
- Free-text keywords
- Language
- License
- Attribution
- Regions
- Data quality statement
- Restrictions
- Other constraints

Other relevant metadata are being added to achieve the best compliance than the possible with INSPIRE directive prescriptions.

Fore remote services some information, such as the abstract, could be inherited directly by the remote source depending on the degree of metadatation of the information source.

Metadata can also be downloaded via the user interface choosing among the two formats: “ISO Metadata” or “Dublin Core Metadata”.

At the present time, all the datasets (both from local or remote source) have a minimum level of metadatation.

Data sharing practices and restriction

Data are shared via the InfoMAPNode platform in view and, where available, in download. Data visualization and download could occur both via graphical interface (the viewer and the appropriate button “download” of InfoMAPNode) or as remote service. In fact, InfoMAPNode has activated WMS, WFS, WCS and WPS OGC standards. This allows both to mirror InfoMAPNode on other platforms in turn, and to access directly data via remote service for example through desktop applications (such as QGIS). These sharing practices are well known and agreed among the geographical information international community, and they allow for a general increase of accessibility, interoperability and, ultimately, reuse of data, in compliance with the FAIR data management principles.

InfoMAPNode doesn’t host for the moment sensitive data. However, private data can be shared by the platform administrator among specific users or groups by setting the appropriate access permissions in “Sharing” section. In particular, five levels of action are foreseen for different users:

1. None: the dataset is not visible to the specific user.
2. View: the user can view the dataset.
3. Download: the user can view and download the dataset.
4. Edit: the user can view, download and edit an existing dataset.
5. Manage: the user can view, download, edit and manage (upload and delete) a dataset.

Other products characterization

Aside from geographical layers, InfoMAPNode also handles documents, maps, geostories and dashboards.

Documents are represented not only by textual documents, but also spreadsheets, archives, images, videos and other material are included in the category. Many different file formats are supported (.txt, .log, .doc, .docx, .ods, .odt, .sld, .qml, .xls, .xlsx, .xml, .bm, .bmp, .dwg, .dxf, .fif, .gif, .jpg, .jpe, .jpeg, .png, .tif, .tiff, .pbm, .odp, .ppt, .pptx, .pdf, .tar, .tgz, .rar, .gz, .7z, .zip, .aif, .aifc, .aiff, .au, .mp3, .mpga, .wav, .afl, .avi, .avs, .fli, .mp2, .mp4, .mpg, .ogg, .webm, .3gp, .flv, .vdo, .glb, .pcd, .gltf) and since they are uploaded into InfoMAPNode they can be easily linked to other data. In a certain sense InfoMAPNode could work as a repository for different kinds of data. At the present time, only few documents have been added to the Documents section and it is not forseen to use InfoMAPNode as a repository since the UNEP-MAP library will be presumably handled externally with respect to GeoNode.

Maps are a result of the superimposition of more geographical layers. InfoMAPNode allows the registered users to create maps starting from the layers present in the system and it also allows layers stylization to be modified for the purpose of the map. In a certain sense the system could be used to easily create rapid geographical representations. It is also possible to print the map and set up the quality and dimension of the output.

Geostories are a powerful communication tool able to connect geographical layers, text, media (videos and images), to embed websites, dashboards and other material. At the present time only two geostories have been added to InfoMAPNode to test their functionality, but the number is foreseen to increase since this instrument has a relevant divulgation pontential, especially for “base users” such as citizens wanting to know more about the MAP work and action areas. Moreover, a geostory can also be useful for educational purposes and for all the actions involving Ocean Literacy in general.

Dashboards are instruments able to analyze data that is represented not only by geographical layers but also (for example) by excel files that can be added as documents to InfoMAPNode. They allow for multiple graphic representations, maps and layers embedding and text tools. At the present time only one test dashboard has been created but in the future the usage of this instrument is foreseen to analyze MAP data. In fact, this kind of tool could be useful to satisfy the needs of a more expert public (such as researchers, stakeholders or community of interests) wanting to have a deeper understanding of data.

For what concerns map, geostories and dashboards ownership and licenses, the owner in this case is the creator of the product, and also license to be attached to the product it is supposed to be decided by the owner. Nevertheless, it is necessary to respect the license of the superimposed layers/used material, since the output product can be at all effects considered as a derivative product. An example could be the use of a layer with a CC-BY license. The name of the layer and the owner must be cited in the final product description. Another example could be the use of a layer shared with CC-BY-NC license. In this case, besides the data and owner citation, it is necessary that the product (map, geostory or dashboard) is not used for commercial purposes.

Data metric policy

Workflows for different access levels

Possible users can be identified in two groups: registered users and non-registered users.

Authentication, authorization, and accounting (also called AAA) is the architecture behind the InfoMAP System to manage intelligently controlling access to UN Environment Programme/MAP resources, enforcing policies, and providing the information needed to use for services. These three elements are considered important for effective network management and security.

The three pillars to control security and right of actors are:

- Authentication is the process of ascertaining that somebody really is whom they say they are.
- Authorization refers to rules/permissions that determine who is allowed to do what.
- Accounting is about keeping track of the resources used for financial or auditing purposes.

Authentication technology provides access control for systems by checking if a user's credentials match the credentials in an authorized user database or in data authentication server. Users are identified with a user ID, and authentication is performed when the user provides a correct credential (password) which matches with the user ID in the database. Each authenticated user can access and manage data domains, based on the user's configured role within the System. Each role has a set of corresponding permissions inside the System, in order to manage, edit and view specific data.

The structure of the profiles and their associated rights for registered users in the InfoMAPNode are:

• Contracting Party users: InfoMAPNode is set up to provide free and accessible spatial data infrastructure for CPs, that can use it for National purpose, allowing to create map using spatial information from different sources and supporting decision makers and public authorities. The contracting party ideally feeds the platform with data from its Country so he will be able both to upload and delete his own material from the platform. Conversely, he will not be able to edit or delete other's material. They can access also restricted data (view and download) under specific privileges. Inside a group of users referring to the same CP there could be different access levels after the competence of the user, whose maximum privileges are the ones described above. In 2022 the formal request for identification of users was sent to MAP FPs. Nominated CPs

can be associated to specific restricted groups with restricted access to specific material, if necessary, and each CP can use this space for internal use also.

• MAP Component users: Users which are staff of MAP Secretariat and Component; for each of them, there is a different role in the System due to the competence and role of the activities carried out in the different data flows and data assessment. The subdivision is the following:

  - CU, other RACs and MAP Components: the RACs often participate in data harmonization and rework, so they can upload and delete their data from the platform as happens for CPs. They cannot edit or delete other authors' material, but they can access (view and download) data under higher privileges than the unregistered user.
  - INFO/RAC: INFO/RAC is the administrator of InfoMAP Node. INFO/RAC holds all rights in order to protect data and system security. INFO/RAC is responsible for system administration: it can upload and delete all the data of the platform; it can view and download all the data and set access privileges for each group of registered and non-registered users.

- Stakeholders: stakeholders are a group of or single users that need (for different reasons) to access MAP data. Ideally it is a registered user that cannot upload and edit/delete data from the platform but can view and download also restricted data. Special privileges are set up by the system administrator based on specific necessities.

Non-registered users are identified under the "Anonymous" group: They represent users who are not authenticated, and they have only the possibility to search, view and in some cases download metadata and data publicly available.

For who manages the whole infrastructure (INFO/RAC members), data ingestion can be made in two principal ways: directly from the Geonode user interface (both claiming publishing of self-archived layers and layers from remote services) or from the underlying Geoserver instance. Data stylization in Geonode interface can be passed to the system by exploiting the capabilities of a standard .sld file, in Geoserver the .sld file can alse be uploaded, catalogued as an available style for layers and associated to the uploaded data. Metadata are compiled through the Geonode interface: a wizard guiding the process is available or an advanced editor can be exploited. Data is published at the end of the whole process or by publishing through the completion of metadatation phase using Geonode interface or by simply publishing it on Geoserver (new available layers are automatically fetched, with their style, by the Geonode instance). For administration users, the whole tools from Geonode are accessible, without restrictions.

For those who publish data/services on the InfoMAPNode platform, all the workflow phases described in the previous paragraph are only available through the Geonode interface. No write access to the underlying Geoserver instance is guaranteed.

For other users, metadata editing is allowed on dataset and documents uploaded by the users themselves or layers linked to a specific group.

Interactions according to data granularity and access

Here the matrix reporting interactions between access levels and allowed actions is reported.

	User	Is registered?	Can Upload?	Can Edit/Delete?	Can View?	Can Download?	Can Set Privileges?
Contracting Parties	CP	Yes	Yes	Yes, their data	Yes. Public and restricted material (following sharing regulations)	Yes. Public and restricted material (following sharing regulations)	No
MAP components	MAP CU and RACs (except INFO/RAC)	Yes	Yes	Yes, their data	Yes. Public and restricted material (following sharing regulations)	Yes. Public and restricted material (following sharing regulations)	No
	INFO/RAC	Yes	Yes	Yes, all the material on platform	Yes, all the material on platform	Yes, all the material on platform	Yes
MAP partners	Stakeholders and researchers	Yes	No	No	Yes, Public and restricted material (following sharing regulations)	Yes, Public and restricted material (following sharing regulations)	No
Other users	Anonymous	No	No	No	Yes, only public material	Yes, only public material	No