GEGN.2/2023/48/CRP.48

13 March 2023

English

United Nations Group of Experts on Geographical Names
2023 session
New York, 1 – 5 May 2023
Item 14 of the provisional agenda *
Geographical names data management

Indonesian geographical names information system (SINAR): improving the interoperability and usability of geographical names

Submitted by Indonesia**

Summary:

In Indonesia, geographical names standardization is aimed at providing a comprehensive, accurate and easily accessible geographical names database. Data integration among stakeholders is a significant challenge in creating a national database. The Geospatial Information Agency (Badan Informasi Geospasial) is developing and advancing its geographical names information system, SINAR (Sistem Informasi Nama Rupabumi), to tackle this issue. The main focus of the current development is to integrate the SINAR database as a national database of geographical names and implement the "create once, use many times" principle. Therefore, the Agency has integrated geographical names from various ministries and developed the geographical names application programming interface.

Currently, data integration is conducted with a data-collection process that is partially utilized by the application programming interface used by some ministries. Meanwhile, some ministries are still using traditional data integration by exchanging file-based data. Next, the data collected are processed and integrated into SINAR to carry out the geographical names verification and standardization process. The verified and standardized geographical names in SINAR can be accessed and utilized by stakeholders in various ways.

* * GEGN.2/2023/1

** ** Prepared by Hanif Arafah Mustofa, Septin Mulatsih Rezki, Ahmad Kosasih, and Muhammad of the Geospatial Information Agency.

To improve the interoperability and usability of geographical names among stakeholders, the Agency initiated the development of the geographical names application programming interface in 2021. This interface allows other systems (managed by ministries and local governments) to access and interact with the geographical names data maintained in the SINAR database. However, implementing this interface has been challenging, especially providing one that meets users' needs. Moreover, it needs further improvement, including with regard to the implementation of user roles in the application programming interface system, application programming interface catalogue provision, geocoding improvement, and implementation of geographical name codification to support linked data, metadata provision and Resource Description Framework generation.

Indonesian geographical names information system (SINAR): improving the interoperability and usability of geographical names

Introduction

Geographical names standardization in Indonesia aims at providing a comprehensive, accurate, and easily accessible geographical names database. As a large nation with 38 provinces and 514 regencies/cities, the geographical names standardization process is challenging due to insufficient funding and proficient human resources, especially in data collection. A rough estimation by BIG shows that the collection of geographical names data will require more than four years, under the condition of unlimited budget and employed hundreds of surveyors, to merely conduct field surveys. Nevertheless, based on the coordination conducted by BIG with several government institutions, many government institutions already have data related to geographical names. The Ministry of Public Works and Housing holds data on geographical names related to housing, while the Ministry of Health has data on geographical names pertaining to health infrastructure. The Ministry of Education and Culture manages data on geographical names of educational facilities, and the Ministry of Religious Affairs maintains data on geographical names of worship facilities and religious education. Therefore, integrating geographical names data among government institutions can be a viable solution to accelerate the collection of geographical names in Indonesia.

To support the process of integrating data among government institutions, BIG has been developing and enhancing its Geographical Names Information System, known as SINAR (*Sistem Informasi Nama Rupabumi*). The primary objective of the current development is to integrate the SINAR database as a national database of geographical names and implement the "create once, use many times" principle. Therefore, BIG has integrated geographical names from various ministries and developed a geographical names Application Programming Interface (API).

Geographical names data integration

BIG and several relevant government institutions launched an initiative to integrate geographical names data. Data integration has commenced with several ministries, especially those that have opened and provided data related to/or containing geographical names data. Ministries involved include the

Ministry of Public Works and Housing, the Ministry of Health, the Ministry of Education and Culture, the Ministry of Religious Affairs, and the Ministry of Tourism and Creative Economy. Currently, BIG and stakeholders use two methods to share toponym data: either through partial utilization of an API between stakeholders or through traditional data integration by exchanging file-based data.

1. Traditional data integration

   The exchange of geographical names data with several government institutions is still carried out through traditional methods, which involve file-based data exchange, such as tabular or CSV files. BIG has implemented the method with the Ministry of Tourism and Creative Economy and the Ministry of Religious Affairs. The data integration process includes data identification, mapping matrix preparation, data conversion, and integration into the SINAR database. The initial data identification is conducted to determine the structure and content of the initial data (Fig. 1). Based on the identification, a mapping matrix is prepared if there is a difference in data structure compared to the SINAR data structure (Fig. 2). After mapping matrix preparation, the data conversion process is carried out along with data improvement (Fig. 3). The resulting converted data is then uploaded into the SINAR database in vector format through the administrator module (Fig. 4).

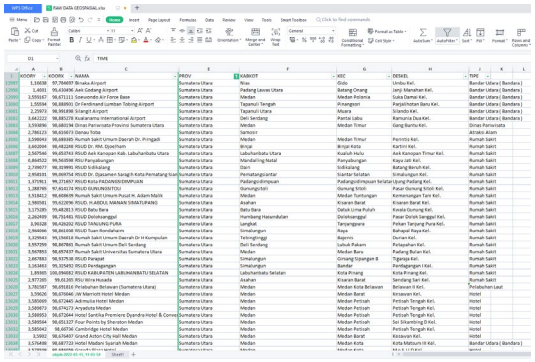

Figure 1. The screenshot of initial geographical names data from Ministries of Tourism and Creative Economy

Struktur Data dari Kementerian Pariwisata		Struktur Data SINAR	
Field Name	**Description**	**Field Name**	**Description**
KOORDY	Longitude	Bujur	Longitude
KOORDX	Latitude	Lintang	Latitude
NAMA	Name of objects	Nama Generik/Lokal	Generic Element
		Nama Spesifik	Specific Element
PROV	Province	Wadmpr	Province
KABKOT	Regency/City	Wadmkk	Regency/City
KEC	District	Wadmkc	District
DESKEL	Village	Wadmkd	Village
TYPE	Feature type	Ftype	Feature type

Figure 2. The screenshot of mapping matrix preparation

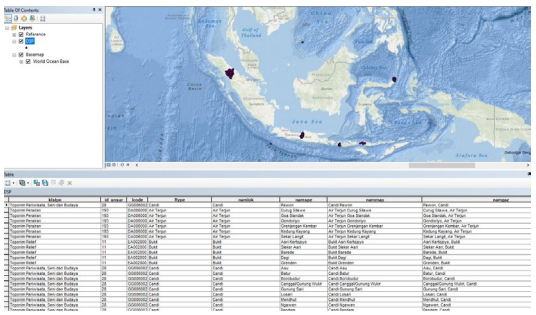

Figure 3. The screenshot of conversion data result

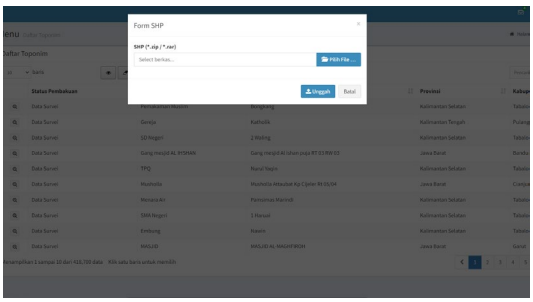

Figure 4. The screenshot of data uploading in SINAR

2. Data integration using API

   In addition to file-based data exchange, several government institutions have data on geographical names through API. As a result, BIG has also started implementing data exchange using API provided by the Ministry of Public Works and Housing and the Ministry of Health. The data

exchange through API is carried out with a mechanism where BIG is granted access rights to geographical names data from the ministry through the API in the form of downloading relevant geographical names data (Fig. 5 and Fig. 6). Using this mechanism, BIG can obtain updated data from the ministry anytime. However, the weakness of this mechanism is that traditional steps are still needed, including identification, matrix mapping (once at the beginning), data conversion, and data uploading to the SINAR database.

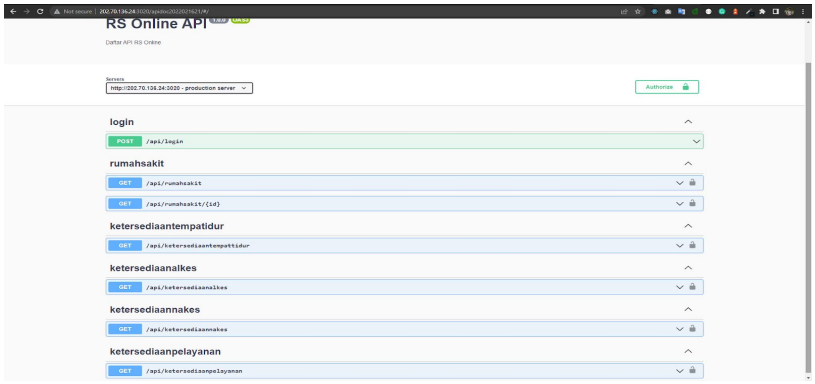

Figure 5. The screenshot of API provided by the Ministry of Health

Figure 6. The screenshot of API provided by the Ministry of Public Works and Housing

Development of the SINAR API

The geographical names data integrated into SINAR are then verified according to the respective authority. The verified data at the district/city, province, and central levels are then publicly announced for 30 working days before being standardized. The verified and standardized geographical names are accurate and can be reused by government agencies and the public. The standardized geographical names can be downloaded in vector file, JavaScript Object Notation (JSON), and tabular formats. However, exchanging data in a file-based format can become complicated when a system needs to access this data. Government agencies need to download it from SINAR, then download it again when there are changes. The file-based data exchange method is less efficient when used by other systems. To improve the interoperability and usability of geographical names among stakeholders, the Agency initiated the development of the geographical names API in 2021.

The geographical names API enables other systems managed by ministries and local governments to access and interact with the geographical names data maintained in the SINAR database. It allows

stakeholders to access the latest and dynamic geographical names data according to their needs without directly accessing the SINAR database. The geographical names API was developed in REST API type that will produce data in JSON format. Initially, the development of API resulted in two endpoints: geographical names and historical data. Each developed endpoint will be explained in the following points:

1. Geographical names data

   Other systems can request data using geographical names API to obtain geographical names alongside the available parameters according to their needs. For example, the user could search for a geographical name in a specific province or feature type by using the "filter_provinsi" and "filter_unsur" parameters. Detailed information about the geographical names endpoint is presented as follows:

Base URL	https://sinarapis.big.go.id/sinar/v1
Endpoint	/v1/toponim
Method	GET
Parameter	• limit • skip • keyword • display_fields • sort_by • filter_provinsi • filter_kab_kota • filter_kecamatan • filter_desa_kelurahan • filter_desa_kelurahan • filter_unsur • filter_id_toponim • filter_nama_surveyor • filter_status_data • filter_status_pembakuan • filter_tanggal_survei • filter_nama_lokal • filter_nama_spesifik • filter_asal_bahasa • filter_arti_nama • download

Figure 7. Detailed information about the geographical names endpoint

If the access to the geographical names data endpoint is successful (status code 200). The geographical names data generated as it exists in the SINAR database such as geographical name, feature type, historical name, meaning of name, and especially for "geometry" attribute, which is processed from the "geom" field and produces part of the GeoJSON format. The example of the geographical names API for Geographical Names Data response present as follows:

```
{
  "success": true,
  "code": "200",
  "message": "OK",
  "data": {
    "rows": [
      {
        "namlok": "Kantor Bupati",
        "namspe": "Bekasi",
        "id_toponim": 442388,
        "geom": "0101000020E6100000B56035D80CCB5A40AD889AE8F37519C0",
        "id_pengguna": 8247,
        "id_unsur": 273,
        "negara": "INDONESIA",
        "id_provinsi": "32",
        "id_kabkota": "32.16",
        "id_kecamatan": "32.16.20",
        "id_desa_kelurahan": "32.16.20.2002",
        "id_lingkup": 4,
        "id_status_pembakuan": 1,
        "id_status_data": 1,
        "id_status_publikasi": 1,
        "nlp": "",
        "alias": "",
        "aslbhs": "",
        "artinam": "",
        "sjhnam": "",
        "nambef": "",
        "namrec": "",
        "elevasi": "80.9981",
        "foto1": "https://sinar.big.go.id/assets/images/foto/442388_1.jpg",
        "foto2": "https://sinar.big.go.id/assets/images/foto/blank.png",
        "foto3": "https://sinar.big.go.id/assets/images/foto/blank.png",
        "foto4": "https://sinar.big.go.id/assets/images/foto/blank.png",
        "rekaman1": "https://sinar.big.go.id/assets/audio/442388_1.mp3",
        "rekaman2": "",
        "narsum": "",
        "tglsurvei": "2023-02-28",
        "akurasi": "3.79009",
        "sumber": "Data Hasil Survei Toponim",
        "remark": "",
        "sketsa": "https://sinar.big.go.id/assets/images/sketsa/blank.png",
        "ucapan": "",
        "ejaan": "",
        "created_date": "2023-03-06 08:27:45.493206",
        "created_by": 8247,
        "updated_date": "2023-03-06 08:27:45.67882",
        "pendataan_type_id": 1,
        "pendataan_nonsi_id": 0,
        "dokumen1": "https://sinar.big.go.id/assets/document/blank.png",
        "jenis_toponim": "Titik",
        "zonautm": "48S",
        "id_klasifikasi_toponim": 21,
        "icode": "GD002022",
        "nammap": "Kantor Bupati Bekasi",
        "namgaz": "Bekasi, Kantor Bupati",
        "koordinat1": "06°21'54.67\" S  107°10'21.57\" T",
        "koordx": "107.17265897",
        "koordy": "-6.36518825",
        "nsurveyor": "Desi Puspitasari",
        "minx": "107.17265897",
        "miny": "-6.36518825",
        "maxx": "107.17265897",
        "maxy": "-6.36518825",
        "centroidx": "107.17265897",
        "centroidy": "-6.36518825",
        "geometry": {
          "type": "Point",
          "coordinates": [
            107.17265897,
            -6.36518825
          ]
        },
        "b_box": "0101000020E6100000B56035D80CCB5A40AD889AE8F37519C0",
        "id_statpub": 1,
        "statpub": "Terbit",
        "id_statpem": 1,
        "statpem": "Data Survei",
        "id_status": 1,
        "status": "Proses",
        "id_klstpn": 21,
        "klstpn": "Toponim Pemerintahan",
        "ftype": "Kantor Bupati",
        "lingkup": "Desa / Kelurahan",
        "id_wilayah_administrasi": "32.16.20.2002",
        "wadmpr": "Jawa Barat",
        "wadmkk": "Bekasi",
        "wadmkc": "Cikarang Pusat",
        "wadmkd": "Sukamahi",
        "ksurveyor": 8247
      }...]
    }
}
```


Figure 8. The example of geographical names API for Geographical Names Data response code 200

2. Historical data of geographical names

In addition to the geographical names data endpoint, other systems can also obtain information related to the historical processing of geographical names data, including information on when the data was entered into the system and when it was verified at the district/city, provincial, and central levels. Detailed information about the historical geographical names endpoint is presented, and an example response is provided as follows:

Base URL	https://sinarapis.big.go.id/sinar/v1
Endpoint	/v1/toponim/:id/riwayat-pembakuan
Method	GET
Parameter	• limit • skip • sort_by

Figure 9. Detailed information about the historical geographical names endpoint

```
{
  "success": true,
  "code": "200",
  "message": "OK",
  "data": {
    "rows": [
      {
        "id_riwayat_pembakuan": 377960,
        "tanggal": "2022-01-10 14:46:14.973085",
        "keterangan": "",
        "status": "Terima",
        "statpem": "Penetapan"
      },
      {
        "id_riwayat_pembakuan": 48142,
        "tanggal": "2020-09-23 22:39:29.825812",
        "keterangan": "",
        "status": "Terima",
        "statpem": "Penelaahan Pusat"
      }
    ],
    "metadata": {
      "count": 2,
      "limit": 10,
      "sort_by": "oldest"
    }
  }
}
```


Figure 10. The example of geographical names API for historical geographical names endpoint Response code 200

Further developments

Implementing geographical names API has been challenging, especially providing one that meets users' needs. Moreover, it needs further improvement, concerning the implementation of user roles in the application programming interface system, API catalogue provision, geocoding improvement, and implementation of geographical names codification to support linked data, metadata provision, and Resource Description Framework (RDF) generation.

1. API improvement based on users' needs

   The need for API in the systems owned by stakeholders is increasing, indicated by the number of systems built by stakeholders that have data related to geographical names. For example, SIGI PUPR (https://sigi.pu.go.id/ast/) is a WebGIS that has features to visualize infrastructure in Indonesia that has been built by The Ministry of Public Works and Housing. Satu Peta Jabar (https://satupeta.jabarprov.go.id/#/map) is a web portal that connects geospatial data-related systems in West Java Province. *Wisata Tangguh* Kemenparekraf (https://wisatatangguh.kemenparekraf.go.id/)is a website developed by the Ministry of Tourism and Creative Economy for visualizing data related to tourism.

   Based on several recent meetings with stakeholders, it can be concluded that stakeholders need an API from the SINAR database to enhance their systems to match the geographical names in SINAR. However, more than the existing condition of geographical names, API is needed to meet the demands of stakeholders' systems. For example, The Ministry of Public Works and Housing needs spatial data types available in SINAR. However, the existing API condition only displays data attributes and cannot download spatial data. Although the geographical names data endpoint has a geometry column, it cannot accommodate the full GeoJSON format, which can be loaded into various GIS applications.

   In addition to the need for spatial data, Jabar Digital Service needs historical data processing based on the classification of feature types. The existing API condition only accommodates the history of each data, not a group of data with the same geographical names or feature types classification.

2. The implementation of user roles

The increasing demand for geographical names API across various sectors and demographics has triggered the need to develop and customize features currently unavailable. Strong coordination between BIG and relevant stakeholders is necessary to develop these unavailable API features, to ensure that stakeholders can integrate their systems well and according to their needs.

One of the essential API features that BIG needs to provide is management and role-based access controls to the API to suit user needs. With this feature, user access to specific data can be regulated and limited, thereby enhancing data access security and simplifying data access according to user needs (not displaying all data, which could slow down the access process). Furthermore, the implementation of management and role-based access controls would facilitate API administrators to monitor stakeholder usage of the API.

For example, implementing management and role-based access controls would enable the Ministry of Tourism and Creative Economy to access data exclusively based on its needs and prevent access by other stakeholders. This development would have significant implications for improving the efficiency and efficacy of the geographical names API and furthering the goal of sustainable and inclusive development.

3. API catalogue provision

API Catalogue is a complete list or collection of APIs available and used by application developers or programmers. Providing an API Catalogue in the geographical names API can help developers of other systems in developing applications, especially in the API integration process, ease of accessing the geographical names API, and access to the geographical names API documentation.

4. Geocoding improvement

The essential utilization of geographical names in this era is utilizing location search processes on maps. Currently, in providing geographical map services to users, BIG offers map services on the tanahair.indonesia.go.id website. The location search feature available on this service does not yet use verified and established geographical names data (still using third-party geocoding services), which may not guarantee data quality.

On the other hand, the quality of data from verified and standardized geographical names data can be accounted for. The quantity of verified and standardized geographical names has also been increasing (a total of 27.910), in line with the increasing activities of geographical names administration. Therefore, the need for geocoding development from verified and standardized geographical names data is essential. The results of this geocoding development can then be implemented in base map services to support location search processes and also be implemented in other systems.

5. Implementation geographical names codification to support linked data

Implementation of geographical names codification refers to the process of assigning unique identifiers or codes to geographical names or locations. This enables the integration and linkage of data from different sources and facilitates interoperability between systems. By adopting a standardized geographical names codification system, geographical names can better support linked data, which is a key aspect of the semantic web. Linked data is a way of connecting data sets and making them available for reuse and sharing across multiple applications and platforms. For example, the Ministry of Tourism and Creative Economy also needs the linkage of data between the SINAR database and the *Wisata Tangguh* (resilient tourism) database. In case there are changes in the SINAR database, the data in *Wisata Tangguh* will also change. However, this requires an additional unique ID column that can link between the two databases, which is not yet available in the existing API.

The use of geographical names codification in SINAR API would enable the accurate representation of location-based data and facilitate the exchange of data with other systems. It would also enhance the ability of geographical names API to support metadata provision and RDF generation, as these standards rely on unique identifiers to link data.

Overall, the implementation of geographical names codification to support linked data would enhance the interoperability, accuracy, and efficiency of geographical names API, thereby contributing to sustainable and inclusive development.

6. Metadata provision

The geographical names metadata should accompany the main data exchange. Metadata is information in a standardized structure and format that describes data, explains data, and facilitates searching, using, and managing data information (Presidential Regulation no. 39 of 2019). The applicable spatial metadata standard in Indonesia is SNI 8843-1:2019, and the metadata service is the Content Services Web (CSW) under SE BIG no. 6 of 2021.

Currently, the downloaded or exchanged data through the geographical names API or SINAR is not equipped with metadata. However, in 2021, BIG initiated the concept of compiling metadata for geographical names in the form of one geographical name, one metadata, considering the unique characteristics of geographical names. This metadata can be used to view the history and quality of geographical names data. Therefore, in the future, BIG needs to focus on developing SINAR to provide metadata, especially related to the automatic generation of metadata features in each feature class.

Other metadata-related developments that can be done include the development of APIs equipped with features that allow users to access or download metadata when accessing geographical names data directly.

7. Resource description framework generation

RDF-based search is a method of information retrieval that utilizes the RDF as the foundation for searching information. This method allows users to perform more specific and sophisticated searches by utilizing information related to the objects contained in the RDF. The utilization of RDF in the toponymy field is crucial, especially in location-based searches that require more specificity. RDF-based search differs from query-based searches in that it is more limited in nature, for example, a location-based search based on queries that can include searching based on geographical names/administrative regions/feature types. An example of a query-based toponym search is "Find toponyms located in Bogor Regency." Meanwhile, an RDF-based toponym search can connect geographical names data with other data, allowing for detailed searches such as "Find places of worship located near the Hermitage Hotel" or "Find road names named after heroes in Indonesia" to be provided to users.

Points for discussion

The Group of Experts is invited to:

1. Express its views on the report and discuss the future development of technology on the interoperability and usability of geographical names data.