D4.1: Open data sets for ML-based RRM

Revision: v1.0

Work package	WP 4
Task	Task 4.3
Due date	31/01/2024
Submission date	05/02/2024
Deliverable lead	HSP
Version	V1.0
Authors	Enrique Marin (HSP), Nuria Trujillo (HSP), Fanny Parzysz (ORA), Laurent Reynaud (ORA), Marius Caus (CTTC), Luis Blanco (CTTC), Cristian J. Vaca-Rubio (CTTC)
Reviewers	Mathieu Arnaud (TAS-F)
Abstract	Dataset to be used during 5G-STARDUST project within WP4 and possibly also within WP5
Keywords	Database, satellite, RRM, 5G

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1		First draft: ToC definition	Marius Caus (CTTC)
V0.2	28/12/2023	First draft “Satellite Network Dataset” section	Enrique Marín (HSP) & Nuria Trujillo (HSP)
V0.2.1	09/01/2024	Minor modifications “Satellite Network	Enrique Marín (HSP) & Nuria Trujillo

Co-funded by the European Union 6GSNS

Grant Agreement No.: 101096573 Topic: HORIZON-JU-SNS-2022-STREAM-A-01-02
Call: HORIZON-JU-SNS-2022 Type of action: HORIZON-JU-RIA

		Dataset" section	(HSP)
V0.3	12/01/2024	First draft "Cellular Network Dataset" section	Fanny Parzysz (ORA)
V0.4	15/01/2023	Consolidated "Satellite Network Dataset" section	Enrique Marín (HSP) & Nuria Trujillo (HSP)
V0.5	18/01/2023	Final Version "Satellite Network Dataset" section	Enrique Marín (HSP) & Nuria Trujillo (HSP)
V0.6	29/01/2023	Consolidated Introduction & Purpose of this dataset section	Marius Caus, Cristian J. Vaca-Rubio, Luis Blanco (CTTC) & Alessandro Guidotti (CNIT)
V0.6.1	29/01/2023	Section 2	Fanny Parzysz (ORA)
V0.7	30/01/2024	Inputs for various sections	Marius Caus (CTTC)
V0.8	02/02/2024	QA review from TAS-F	Mathieu Arnaud (TAS-F)
V1.0.F	05/02/2024	Final review approved for upload onto EC portal	Tomaso de Cola (DLR)

DISCLAIMER

5G-STARDUST (*Satellite and Terrestrial Access for Distributed, Ubiquitous, and Smart Telecommunications*) project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096573.

COPYRIGHT NOTICE

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	**R**	
Dissemination Level		
PU	*Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)*	✓
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	

Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

EXECUTIVE SUMMARY

This report has been prepared in the framework of WP4 (Radio Technologies) and specifically within the scope of Task 4.2 “User-Centric and Digital Beamforming Solutions” and Task 4.3 “Data-enhanced Radio Intelligent Controller Design”. Deliverable D4.1 describes the data sets that will contribute to the design of the beamforming and to efficiently manage the resources of TN and NTN.

The data sets have been divided into two parts, i.e., cellular and satellite network data sets. The document details the features and the format of the data collected in both networks. The Section 2 describes the data set that is the result of monitoring a cellular network operated by ORANGE in France. The inputs of the cellular data set are related to the rural and the railway environments. The data set detailed in Section 3 is obtained from a satellite network operated by HISPASAT that targets America and Europe. The inputs provided by the satellite operator are related to broadband services for fixed clients (residential broadband) and for mobile communications (maritime communications, in this case).

The last part of the document identifies in Section 4 possible applications of the data sets. To this end, Section 4 analyses the role of AI/ML techniques to enhance the performance of an integrated TN-NTN. In the context of Task 4.2, the data sets can be exploited to simplify the beamforming design, by making channel predictions and reducing the complexity of NP-hard problems. In the context of Task 4.3, the data sets can be harnessed to distribute the load between TN and NTN and to dynamically schedule users and allocate resources.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

5G-STARDUST	5G-Satellite and Terrestrial Access for Distributed, Ubiquitous, and Smart Telecommunications
AI/ML	Artificial Intelligence/Machine Learning
AMZ-5	Amazonas-5 Satellite
CN0	Carrier-to-noise density ratio
CSV	Comma-Separated Values
EIRP	Effective Isotropic Radiated Power
E/S	Earth Station
FEC	Forward Error Correction
GEO	Geosynchronous Earth Orbit
G/T	Gain-to-Noise-temperature
HPA	High-Power Amplifier
H30W-6	Hispasat 30W-6 Satellite
IBO	Input Back-Off
KML	Keyhole Markup Language
KMZ	Keyhole Markup Language Zipped
MODCOD	Modulation and Coding
NTN	Non-Terrestrial Networks
OBO	Output Back-Off
PDF	Portable Document Format
PEB	Power Equivalent Bandwidth
RRM	Radio Resource Management
SFD	Saturation Flux Density
SINR	Signal-to-Interference-Plus-Noise Ratio
VBA	Visual Basic for Applications
WP	Work Package

1 INTRODUCTION

1.1 SCOPE

This deliverable D4.1 provides close to real operational data from a satellite network and a cellular network. Building upon the data collected in D4.1, 5G-STARDUST will investigate the applicability of artificial intelligence (AI) and machine learning (ML) techniques for designing the beamforming and performing radio resource management optimizations and predictions in an integrated terrestrial and non-terrestrial network. The AI/ML-based developments will be carried out in Task 4.2 “User-Centric and Digital Beamforming Solutions” and Task 4.3 “Data-enhanced Radio Intelligent Controller Design”. The deliverable D4.1 sets the basis to fulfil one of the objectives of the project, which is the design of data-driven management system components, building on AI/ML-based solutions for resource allocation and service provision in highly dynamic integrated hybrid networks.

1.2 AI-ENHANCED TN-NTN

Integrating AI into terrestrial networks (TN) and non-terrestrial networks (NTN) offers several advantages. For instance, AI enhances dynamic resource allocation, optimizing data transmission routes between either satellite and terrestrial nodes based on real-time conditions. This leads to overall enhancement in network efficiency by minimizing latency and/or maximizing bandwidth usage.

What is more, AI-driven predictive analysis can forecast potential network congestion, enabling proactive measures to prevent disruptions. This forecasting capability ensures a more reliable communication infrastructure, being beneficial for applications such as telemetry, critical data transmission or congestion control.

Additionally, AI can enhance automation for fault detection. By continuously monitoring network performance and identifying anomalies, AI algorithms can trigger responses to mitigate issues before they broadly escalate, minimizing downtime and making an efficient use of routing via satellite or terrestrial networks.

Furthermore, the inclusion of AI in a joint satellite-terrestrial network potentially improves the adaptation capabilities of the whole network infrastructure. As consequence, a more reliable, robust, and intelligent network can meet the diverse demands of modern satellite-based applications.

1.3 ORGANIZATION OF THE DOCUMENT

The data has been gathered from a cellular network operated by ORANGE in France and from two satellites operated by HISPASAT, which are oriented to America and Europe. More detailed information is included in the following sections, which are organized as follows:

Section 2 for the Cellular network dataset Includes the general information about the dataset proposed, including the data collection tool and the general information data for the terrestrial part. In addition, Section 2 contains a sub-section concerning the description of the two scenarios chosen for the provided inputs, in this case related to the rural environment and the railway scenario.

Section 3 for the Satellite network dataset Includes all information related to the radio parameters for the satellite inputs studied in the 5G-STARDUST project. This section is divided into four different sub-sections.

- Information about the service provided: overview on the services selected as inputs for the development of the 5G-STARDUST RRM.
- Geographical information: general information about the satellites, oriented to America and Europe, analysed to extract the inputs for the project.
- Information about user and relative position: section related to the relative position of the end-users in the residential broadband scenario.
- Satellite Radio parameters: Section oriented to the satellite radio parameters, indicating what information the dataset contains.

Section 4 for the purposes of this dataset contains an overview of the next steps for the development of the RRM and how these inputs, in coordination with WP5, are useful for its implementation.

2 CELLULAR NETWORK DATASET (ORA)

2.1 GENERAL INFORMATION ABOUT THE PROPOSED DATASET

The cellular network dataset is based on two distinct environments: the rural scenario and the railway one. For both, the weekly traffic pattern of typical base stations is deduced from the monitoring data of 4G commercial networks operated by Orange, in France. Note that, the 5G market has not yet reached its full capabilities (in terms of deployment and user adoption), such that the related data cannot be considered sufficiently representative to be used for the design of TN / NTN resource management mechanisms.

In the following, we describe the main characteristics of the monitoring tool we used for data collection on operational 4G networks. Then, we explain how this raw data has been processed to obtain a valid dataset, for both scenarios.

2.1.1 Tool for data collection on the operational cellular network

The cellular network datasets have been created out of the inputs extracted from Orange's internal tool. This tool allows the monitoring of 2G / 3G / 4G / 5G networks[1] over the whole French territory, thanks to probes at each base station and collecting a wide range of KPIs.

A per-BS data collection: We point out that the proposed datasets have not been elaborated out of drive tests. As consequence, they characterize the traffic load and quality of service aggregated over served UEs. Indeed, to ensure full GDPR compliance, no information related to individual users (identity, localization, mobility or data consumption habits) is collected by this tool.

Time and space granularity: Probes allow the collection of KPIs averaged per week, day, hour and quarter-hour, which could represent a huge volume of data if logged at the lowest time granularity, over a wide geographical area and long period of time. To produce the 5G-Stardust dataset, we could extract hourly raw data over more than one month, and over about one week for the finest time granularity (15min).

Next, eNB-centric metrics, i.e. which aggregate the traffic of all UEs served by a given eNB, would naturally infer geographical metrics, such as a density of connected UEs or a network resource consumption per m^2. However, precautions need to be taken with such usage. Indeed, the unit area is here equal to the cell coverage, which is a priori not known (e.g. load balancing mechanisms could imply that a UE is not served by its closest eNB) and which may vary over time (e.g. configuration changes or cell switch-off during low traffic hours).

About data sensitivity: The KPIs extracted from eNBs operating on commercial networks are obviously very sensitive, as they could give strategical information, for example about the number of clients and their traffic requests, the network dimensioning and provided QoS, as well as hot spots geolocalization. That's why the raw data, extracted from Orange's internal tool, cannot be published as is. Instead, we have processed this raw data to create fictive but

[1] Other types of tools (different from drive tests) could have been used to obtain a UE-centric dataset, i.e. a dataset representing the typical traffic characteristics of individual UEs (pedestrians, passengers of cars or trains, etc.), as well as the UE density and mobility patterns during a day or a week. However, significant processing and adequate cloud services are required to make data suitable for exploitation and comply with GDPR, which is not aligned with the time and effort planning of the project, whereby this option has been dropped.

representative eNBs that illustrate main traffic characteristics, for both the rural and the railways scenarios.

Note that the proposed datasets can be complemented by other public datasets, in particular "Cartoradio", which is maintained by the ANFR (the French national frequency agency) and provides the main characteristics (including site localization) of all BSs, for the four MNOs operating in France. Cartoradio dataset is available here.

2.1.2 Format of the proposed datasets

For both scenarios, the dataset consists in an Excel file where each line corresponds to one cell observed at a given instant. Figure 1 gives an example and columns can be described as follows:

- *weekday:* the text indicates the day of observation ("Monday" to "Sunday") of the representative eNB identified by the site_code. Without surprise, user traffic is generally following regular patterns, due to their behaviour (e.g. commuting, working, shopping, sleeping, etc…). That's why we provided the view of a full week for each representative eNB.

- *hour:* the hour of the observation is logged using the following format: hh:mm:ss.

- *RRC_Connected_Users_Average:* this KPI shows the average number of UEs having at least one signalling radio bearer during the measurement period (one hour or 15min). A UE that connects several times during this measurement period is thus counted several times.

- *PRB_Load_DL:* it refers to the downlink load in terms of PRB (physical resource block, covering the pair of resource blocks of two consecutive slots in a subframe). It includes user data only and not eNB signalization (based on PDSCH or DTCH depending on the vendor).

- *PRB_Load_UL:* Same as above, but for the uplink.

- *Average_Reported_CQI:* it refers to the CQI reported by each served user and averaged over the observation period (1h or 15min).

This list of KPIs has been consolidated through discussions with T4.1 contributors. It accounts for the dataset envisaged utilization, but also on the capabilities of the monitoring platform, as well as access rights. Specific attention has been given to selecting the right KPIs out of the hundreds available ones, and to ensure data integrity. A few KPIs, initially planned for this dataset, had to be removed due to too many extraction inconsistencies. Also note that KPIs are computed based on built-in counters and are vendor-dependant. Definition alignment was thus required to provide a consistent dataset. For some regions, additional KPIs were available, and have been added in the corresponding dataset:

- *Average_RSRP (resp. Average_RSRQ):* it refers to the RSRP (resp. RSRQ) experienced by each served user and averaged over the observation period.

site_code	sector_code	weekday	hour	RRC_Connected_Users_Average	PRB_Load_DL	PRB_Load_UL	Average_Reported_CQI	Avg_RSRP	Avg_RSRQ
1	1	Friday	00:00:00	15.123450	30.723660	26.953167	9.069907	-110.336167	-11.869408
1	1	Friday	01:00:00	12.370040	13.909488	14.096670	9.454005	-109.531667	-11.886867
1	1	Friday	02:00:00	11.803660	10.362148	13.743925	9.776327	-108.654000	-10.776033
1	1	Friday	03:00:00	11.386277	4.572816	12.427003	10.018783	-112.555667	-12.243800
1	1	Friday	04:00:00	11.166917	4.243374	13.859922	10.176692	-112.034000	-12.082635
1	1	Friday	05:00:00	11.993697	7.602060	16.197157	9.791313	-111.947167	-11.170765
1	1	Friday	06:00:00	12.687125	11.698865	28.531500	9.701960	-113.582000	-12.106900
1	1	Friday	07:00:00	18.623800	18.384283	31.085000	9.250967	-113.884833	-14.087783
1	1	Friday	08:00:00	22.148750	22.014600	47.476050	9.047950	-115.399000	-14.238533
1	1	Friday	09:00:00	23.374783	21.431233	56.736000	9.129167	-116.020667	-14.353833

Figure 1: Example of proposed datasets.

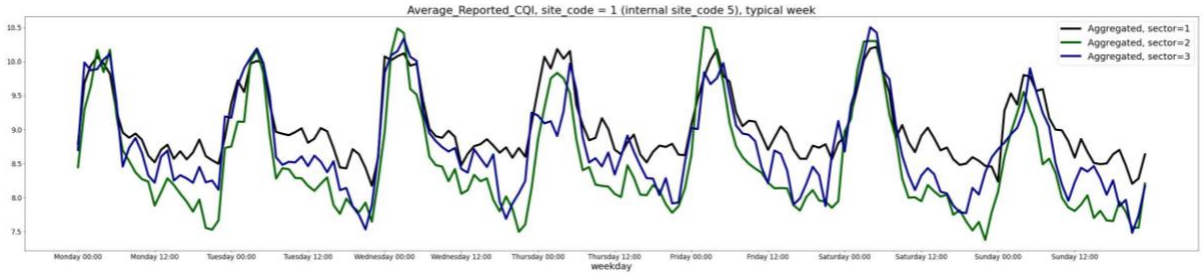

Figure 2: Example of the CQI for eNB labelled 1 in the rural scenario.

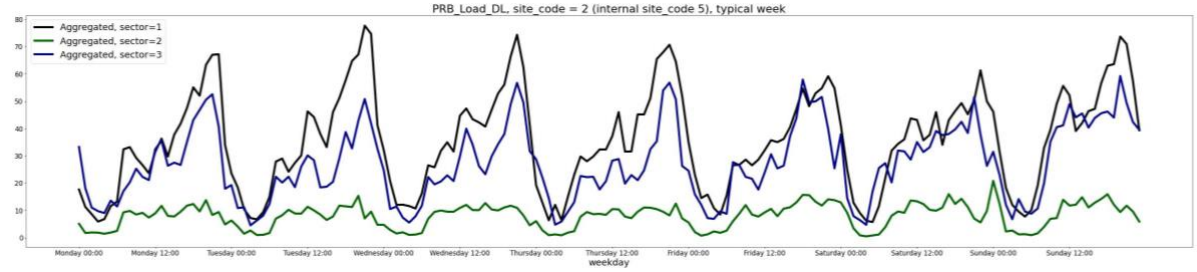

Figure 3: Example of the PRB load (%) for eNB labelled 2 in the rural scenario.

2.2 DESCRIPTION OF TH TWO CONSIDERED SCENARIOS

We now describe each scenario and give details on their geographical characteristics. Specificities related to raw data processing are provided.

Remark: The proposed datasets have been elaborated based on the discussions we had at the time of writing this deliverable. As the activities which will use them have not started yet, they may not show the best fit for future needs. However, it can be envisaged to improve these datasets if required.

2.2.1 Rural environment

This scenario focuses on eNBs located in an environment considered as rural, for example Massif Central, Bourgogne or the centre of Bretagne, as illustrated in Figure 4. It relates to territories with low / very low density of population over vast areas.

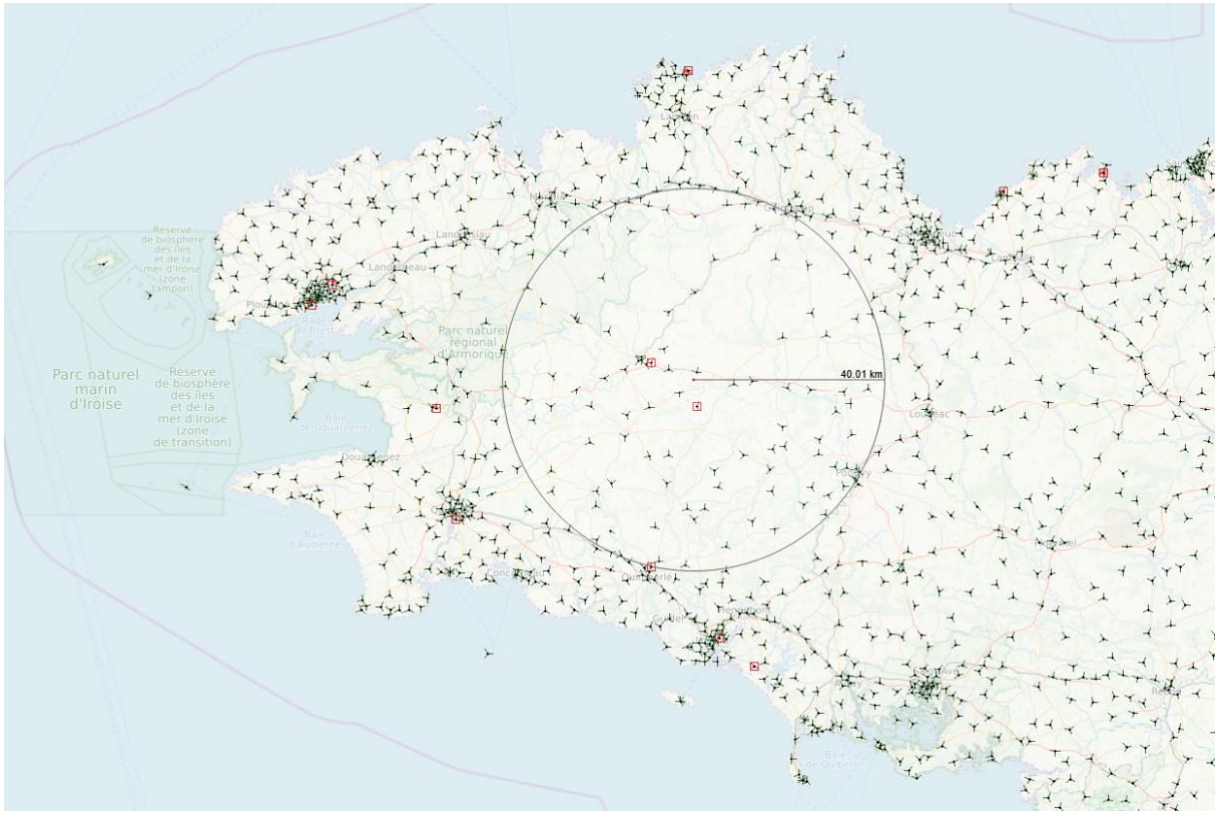

Figure 4: Example of eNBs serving a rural area in Bretagne, France. The circle represents a satellite footprint with a radius of 40km. (Note that this figure does not include recently added stations –

It must be however highlighted that the traffic characteristics extracted from eNB located in these areas **cannot** be associated with residential broadband. Indeed, in France where data has been collected, such type of traffic is conveyed over fixed networks, in particular ADSL, VDSL or optical fibre.

To elaborate the proposed datasets, we considered all eNBs located within circles of 40km-radius (similar to satellite footprint) and located in four different rural regions in France. This represents a total of 422 eNBs. Each eNB is composed of several sectors (usually 3), each with several frequency bands. Several weeks of hourly data have been collected, between October and December 2023, such that a total of nearly 2 million lines have been processed to generate 20 fictive but representative eNBs for the rural scenario.

The final dataset, composed of four different csv files, proposes an average view of rural eNBs, and a particular attention was paid to avoid smoothing of peak hours. Each csv file corresponds to one region, with 5 representative eNBs, 3 sectors for each and 24h x 7days of data, ending into 2520 lines for each csv file. The site identifier consists of a pair of numbers (site_code in {1,5}, sector_code in {1,3}).

2.2.2 Railways scenario

This scenario illustrates the typical traffic pattern of train passengers. The serving eNBs are usually regularly spread along the railroads and consists in two opposite sectors, each pointing to one side of the railroad (there may be additional sectors serving users in the vicinity, but there are not considered in this scenario). Figure 5 illustrates these aspects for the railway between Bordeaux and Poitiers. Note that we focused on railways for high-speed trains (the so-called TGV) and did not consider local railroads. In addition, we have targeted eNBs serving moving trains (out of cities) and not eNBs in the vicinity of main train stations. Indeed, train stations are usually provided with enhanced connectivity capabilities, to absorb the high temporary traffic peaks which occur each time a train arrives or leaves. Hence, the traffic profile differs significantly, and satellite offloading is not expected in this case.

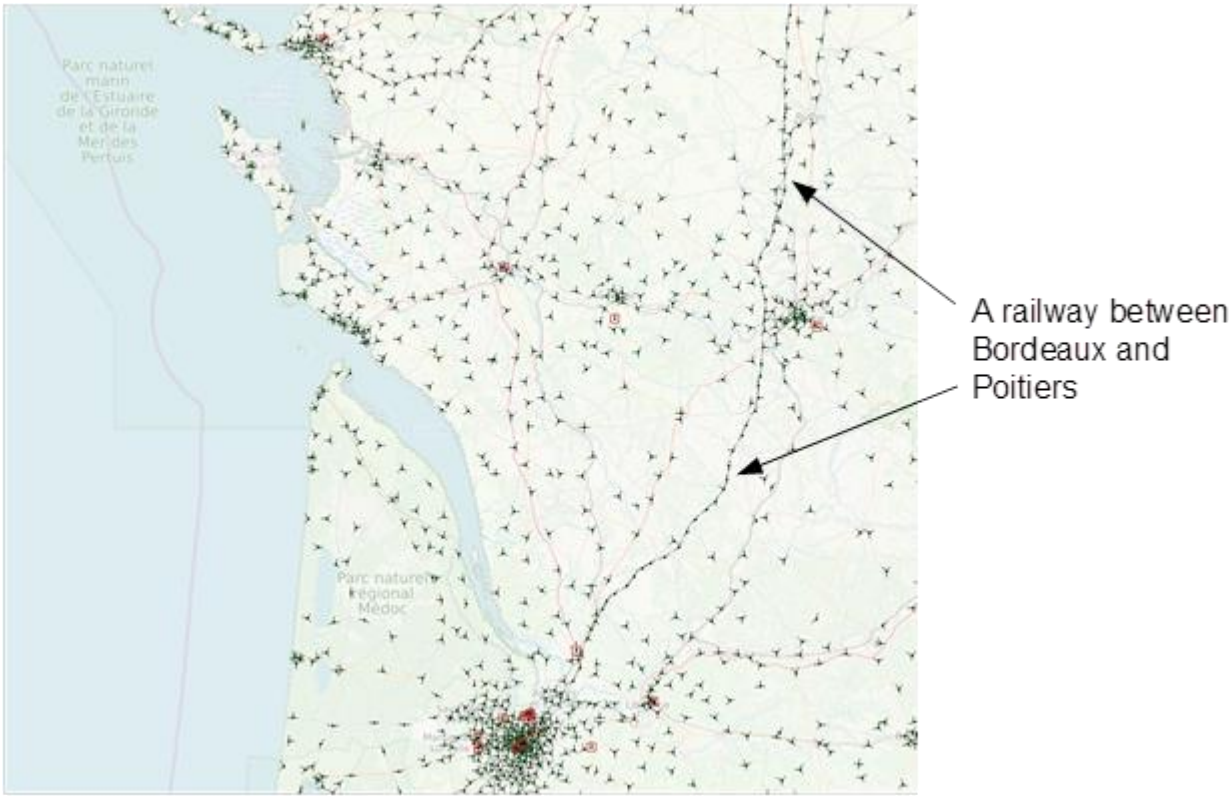

Figure 5: Example of eNBs serving train passengers along a railway (Note that this figure does not include recently added stations – see Cartoradio for up-to-date list)

As for the previous scenario, the proposed dataset aims to reflect the weekly traffic pattern observed at fictive but representative eNBs. In this case, the trains moving along the railways at high velocity imply that a correlation may exist between the traffic observed at a given eNB and at its neighbours. That's why the dataset represents the characteristics of 15[2] consecutive eNBs, as depicted in Figure 6.

[2] The actual number of eNBs will be fixed during the actual simulations' campaigns making use of these datasets within WP4 related tasks (i.e. T4.3).

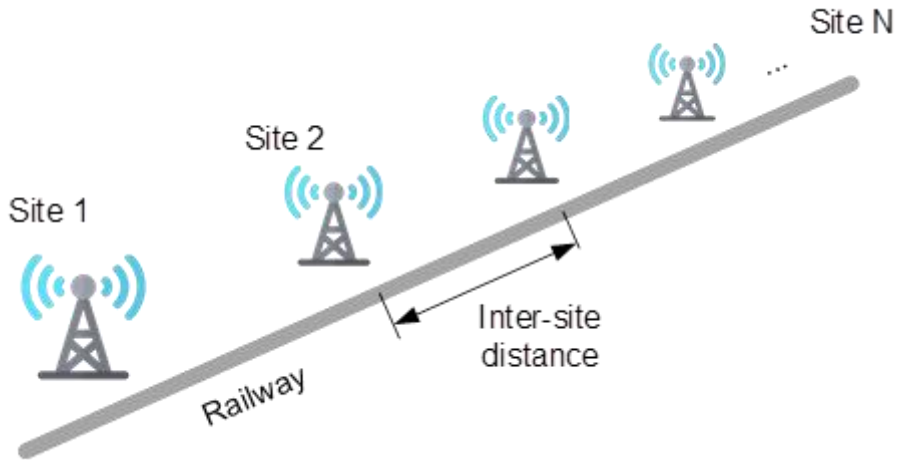

Figure 6: Illustration of consecutive eNBs (fictive but representative) considered for the dataset.

To elaborate this dataset, several weeks of hourly data have been collected, between November 2023 and early-January 2024, such that a total of 150310 lines have been processed.

3 SATELLITE NETWORK DATASET (HSP)

3.1 INFORMATION ABOUT THE SERVICE PROVIDED

The services selected for the elaboration of the satellite network dataset are aligned with the selection of scenarios for the proof-of-concept carried out during 5G-STARDUST WP2.For this purpose, and although the 5G-NTN technology is still in its definition and development, it was defined that the inputs provided by the satellite operator, Hispasat in this case, would be related to the real satellite broadband services for fixed clients (residential broadband) and for mobile communications (maritime communications, in this case).

In order to create a robust and interesting dataset for the project, the satellite operator has defined anonymous clients with a huge number of end users for its elaboration. In this case, the number of clients for residential broadband scenario is more than 500 connected users and for the maritime communications is around 60 ships connected. More precise information is included in the proposed dataset.

The following sections define the geographical information about both indicated services, the satellites used with general information, information about user relative position and satellite radio parameters from the satellite operator's internal platforms.

3.2 GEOGRAPHICAL INFORMATION

In order to provide real information on satellite traffic, we have selected satellites according to the services they offer. Thus, residential broadband and mobility services are not offered by the same satellite or in the same geographical area.

However, both services are transversal in that the inputs come from real values of GEO satellites operated by Hispasat.

In the tables below lists the most relevant characteristics of the satellites that have been selected to extract the inputs for the 5G-STARDUST dataset.

3.2.1 Residential Broadband

The residential broadband scenario comes from real fixed broadband traffic provided in Mexico by the Amazonas 5 Hispasat Satellite (Hispasat, Satellite Characteristics AMAZONAS 5) .The following Table 1 summarizes the general characteristics for the above-mentioned satellite.

Table 1: Amazonas 5 general characteristics (Hispasat, Satellite Characteristics AMAZONAS 5)

Information	Value
Satellite Name	Amazonas 5
Orbital Position	61º West
Launch Date	2017
Transponders	24 Ku Band + 34 Ka band spots

Bandwidth	36 MHz in Ku band + 225 MHz in Ka Band* *Reference value, could vary between different spots
Antennas Number	5
Satellite Type	Transparent (no onboard processing)
Payload Power	10 kW
Available Power	11,5 kW

The AMZ-5 satellite footprint is oriented to America with linear polarization. In this example, the operational transponders that have selected for the datasets are related to Ka spots. The figure below shows the general Ka and Ku beams footprint.

Figure 7: AMZ-5 general America Footprint (Ka (left picture) and Ku (right picture) beams) (Hispasat, Satellite Characteristics AMAZONAS 5)

For this purpose, the datasets involve all the radio parameters information about the beams covered in the Mexico area, the beam size, and the final user position.

3.2.2 Mobility Communications

The mobility scenario comes from real broadband traffic provided in Europe by the H30W-6 Hispasat Satellite. The following Table 2 summarize the general characteristics for the above-mentioned satellite.

Table 2: H30W-6 general characteristics (Hispasat, Satellite Characteristics H30W-6)

Information	Value
Satellite Name	Hispasat 30W-6
Orbital Position	30º West
Launch Date	2018
Transponders	40 Ku Band + 7 Ka band + 10C Band spots
Bandwidth	36 MHz-72MHz in Ku and C band
Antennas Number	5
Satellite Type	Transparent (no onboard processing)
Payload Power	10 kW
Available Power	11,5 kW

The satellite footprint its oriented to Europe and America with linear polarization. In this case, the operational transponders that we have selected for the datasets are related to Ku spots.

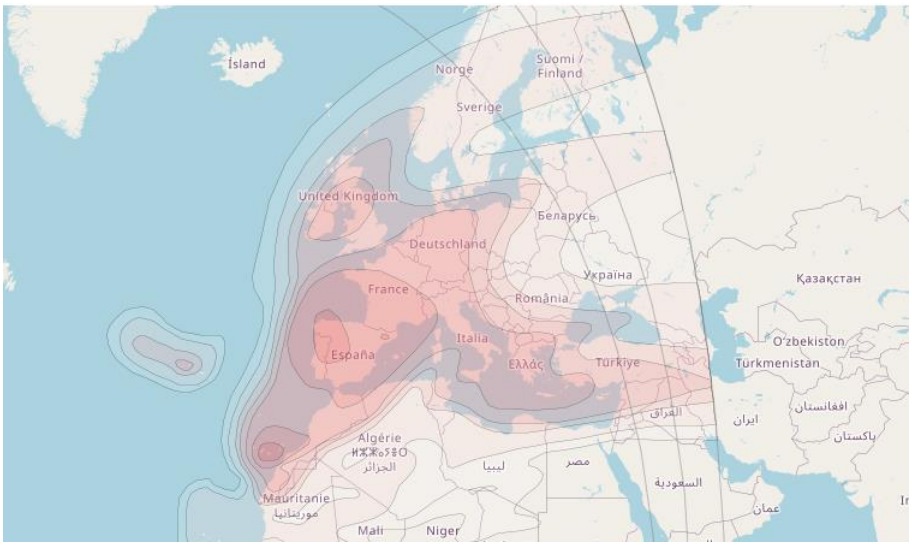

Figure 8: H30W-6 general Europe Footprint (Ku beams) (Hispasat, Satellite Characteristics H30W-6)

Hispasat offers broadband traffic services for maritime communications through the above-mentioned satellite. Since the ships are in motion, and Hispasat only deals with the maintenance of the satellite, its operation and commissioning, but not with its commercialization, we cannot estimate in real time the exact position of the ships.

For this purpose, the data sets include all radio parameter information about the beams covered in the Europe area, the size of the beam and a potential area through which the ships could be moving.

3.3 INFORMATION ABOUT USER AND RELATIVE POSITIONS

The inputs extracted from Hispasat's internal tools collects the real information for a commercial service where the customers are provisioned by satellite beams according to signal parameters, position, or capacity. On the other hand, if a customer is in a location where two beams are overlapped, the satellite operator designates the most appropriate one in terms of Quality-of-Signal and radio parameters.

The inputs extracted from Hispasat's internal tools are in Keyhole Markup Language (KML) and Keyhole Markup Language Zipped (KMZ) format in order to represent the relative UE real position over the real beam footprint.

As has been described before, this information only is available for the residential broadband service because for the maritime communication, since the ships are on the move and Hispasat only deals with the maintenance of the satellite, its operation and commissioning but not its commercialisation, we cannot estimate in real time the real position of the ships.

Table 3: Relative UE Position format and data type

Relative UE Position	Value
Extracted Format	KML/KMZ
Dataset Format	KML/KMZ
Data Type	AMZ-05 beams footprint and Residential Broadband UE position

3.4 SATELLITE RADIO PARAMETERS

Hispasat provides real link budgets with the radio satellite parameters (SINR, CN0, Attenuations…) for the beams covering the geographical locations mentioned above. This link budgets correspond to real information that Hispasat analyses before commercializing a service with the final user and allows Hispasat to select the most appropriated beam for each final client. The commercial link budgets, the same as those proposed for 5G-STARDUST project, are provided in Portable Document Format (PDF) format.

The satellite radio parameters included in the data sets refer to the fields listed in the following tables. As discussed above, this information is the information necessary to size the satellite network ensuring from the engineering side that the satellite link is operating correctly:

Table 4: Space Segment and Earth Station Characteristics fields included in Dataset.

Space Segment Characteristics	Earth Station Characteristics
Uplink Connectivity	Location
Downlink Connectivity	Antenna Diameter (m)
SFD at beam center (dBW/m^2)	E/S HPA (W)

Freq.Band	E/S EIRP @ Saturation (dBW)
Transponder BW (MHz)	E/S G/T (dB/K)
IBO (dB)	Tracking (Y/N)
OBO (dB)	Additional Notes
Txp Operational Mode	

Table 5: UL/DL Calculations fields included in Dataset

Carrier Characteristics – Link Budget results – RTN/FWD
Carrier ID
Uplink
Downlink
Modem Technology
Scenario
Modulation
FEC
Info Rate (Kbps)
Symbol Rate (Ksps)
BW (KHz)
PEB (KHz)
E/S EIRP (dBW)
Threshold Eb/No(dB)
Available Eb/No(dB)
Margin (dB)
Availability (%)

The following link budget has been developed for the 5G-STARDUST consortium:

- Link budget per beam as a reference in the beam peak (AMZ-5)
- Link budget per beam in the beam peak/edge (AMZ-5)
- Link budgets in an area with overlapped beams between beam 38/39 (AMZ-5)
- Link budget in Maritime Zone (H30W-6)

This information must be compared with the radio information extracted from the broadband traffic side, that allow to compare if the studied value is finally achieved or not.

Table 6: Satellite Radio Parameters format and data type

Satellite Radio Parameters	Value
Extracted Format	PDF
Dataset Format	PDF
Data Type	Link budget, radio parameters

4 PURPOSE OF THE DATASET

In the context of User-Centric and Digital Beamforming techniques, to be developed in Task 4.2, the dataset described in Section 3 could provide a significant added value. In general, beamforming techniques can exploit either Channel State Information (CSI) or location estimated measured by the User Equipment (UE) and reported to the network entity computing the radio resource allocation matrix and the beamforming coefficients. In such a case, the satellite network dataset exposed in Section 3 and traffic information provided in deliverable D5.1 "Open Data Sets for AI Data Driven Networking" could be used in a flexible payload scheme to address the next problems:

- AI-based beam selection and/or beamformer design in a non-GEO NTN system. Considering the footprint, the location of the NTN users and the SINRs received by the users in the coverage area of the GEO satellite, some AI-based algorithms can be considered to select/design the beams of a non-GEO system. Different type of problems can be considered (e.g., sum-rate maximization, max-min fair with per antenna power constraints, per-antenna power minimization with QoS constraints, etc.). These problems are in general NP-hard and AI-based algorithms can be considered to reduce the computational complexity of the problems.

It is important to remark that computation is often complex and the performance of the beamforming algorithm is deeply impacted by the latency involved in the process. More specifically, the actual channel that will be encountered by the beamformed transmission will be that at a time t_1, with the resource allocation and the beamforming coefficients computed based on estimates at time t_0. The larger the interval $t_1 - t_0$, the worse the performance. In this context, AI/ML solutions might be extremely beneficial for different objectives, including, but not limited to: i) predictive location/CSI estimation; ii) AI-based beamforming, in which the beamforming coefficients are computed without the need to explicitly elaborate the estimated channel matrix; and iii) AI-based RRM and beamforming, in which both the radio resource allocation and beamforming coefficients are computed through AI solutions. To this aim, in the context of Task 4.3, the different features available in the NTN dataset will be evaluated for these objectives, in particular their correlation or, in any case, relevance in predicting the CSI/locations or computing the RRM/Beamforming matrices.

In the context of T4.3, the objective is to design a RIC performing optimizations devoted to both large scale and small-scale time variations. In the space segment, this includes payload control, satellite beam EIRP and bandwidth allocation among other parameters including UE time-frequency resources scheduling, beam assignment to cells and power control. In this specific context, such optimizations are based upon cost functions (maximizing the delivered capacity, minimizing the overall interferences between the cells, etc.) and this makes the optimization process sensitive to the provided input. For example, the result of such work could vary significantly if the initial repartition of the capacity demand (called traffic model) was considered to be uniform rather than being composed of hotspots. Having a clear overview of the data set provided in D4.1 together with the traffic model in deliverable D5.1 "Open Data Sets for AI Data Driven Networking" is therefore essential to develop the NTN RIC.

Within the scope of T4.3, the data set can also be exploited in an integrated TN-NTN architecture to make an efficient use of the resources. More precisely, with the cellular network datasets exposed in Section 2 the next type of problems can be addressed:

- AI-based offloading strategies for integrated TN-NTN. For this activity both datasets could be potentially merged.

- Predictive modeling and load balancing:

- Use historical data of RRC_Connected_Users_Average, PRB_Load_DL, PRB_Load_UL, and Average_Reported_CQI to predict future values of these KPIs.
- Develop time series forecasting models to predict future values and decide an offloading to the satellite.

- Anomaly detection:
  - Identify abnormal behavior in the network by training an anomaly detection model
    - Unusual spikes/drops in RRC_Connected_Users_Average, PRB_Load_DL, PRB_Load_UL, or Average_Reported_CQI may indicate network issues.
  - Machine learning models might be interesting for flagging this anomaly and deciding using the satellite as a back-up link.
- Resource allocation:
  - Optimize resource allocation based on the predicted future values of PRB_Load_DL, PRB_Load_UL, and Average_Reported_CQI.
- To exploit mobility patterns in the railway system to dynamically allocate resources in the integrated terrestrial and non-terrestrial networks.

5 REFERENCES

Hispasat. (n.d.). *Satellite Characteristics AMAZONAS 5.* Retrieved from https://www.hispasat.com/contenidos/web/0/192-satelite-amazonas-5-de-hispasat-3.pdf

Hispasat. (n.d.). *Satellite Characteristics H30W-6.* Retrieved from https://www.hispasat.com/contenidos/web/0/194-h30w-6.pdf