DESTINI

SMART DATA PROCESSING AND SYSTEMS OF DEEP INSIGHT

http://www.destini2020.eu

Deliverable D1.2

Data Management Plan

Document details:

Editor:	Andreas Andreou
Contributors:	Andreas Andreou, Michael Pingos, Panayiotis Christodoulou, Spyros Loizou, Andreas Christophorou
Date:	19/12/2022
Version:	12.0

Document history:

Version	Date	Contributor	Comments
v.01	06/03/2020	Panayiotis Christodoulou	Initial document structure
v.02	17/03/2020	Andreas Andreou	Review of document structure
v.03	31/03/2020	Panayiotis Christodoulou	First Draft
v.04	09/04/2020	Andreas Andreou	Comments
v.05	13/04/2020	Panayiotis Christodoulou	Corrections
v.06	15/04/2020	Partners	Final Comments
v.07	16/04/2020	Panayiotis Christodoulou	Final Corrections
v.09	17/04/2020	Partners	Approved final version after minor revision
v.10	18/10/2022	Andreas Andreou	Comments for update
v.11	19/10/2022	Panayiotis Christodoulou	Revision & update
v.12	20/10/2022	Partners	Final Approval
V.13	19/12/2022	A.S. Andreou	Document Revision

Contents

1. Introduction

The aim of the DESTINI's Data Management Plan (DMP) is to identify the project's research data and to describe how to make them findable, accessible, interoperable and re-usable (FAIR) following the H2020 Data Management Plan Template v1.0 available on the H2020 EU Portal. This deliverable outlines how research data will be handled during the research project, and after finalization.

Partners of the consortium involved in research activities were asked to provide detailed information to describe what data, methodologies and standards will be used, whether and how this data will be shared and/or made open, how data will be curated and preserved and how it will be collected and/or generated during the whole duration of the project.

Data reported in this deliverable will be made available and interoperable to all stakeholders. To avoid issues related to IP rights and their access, only data related to scientific publications produced during the duration of the project will be made available to the public. DESTINI's Data Management Plan will be a living document and updates will occur over the course of the project whenever significant changes arise such as (but not limited to): new data, changes in consortium policies (e.g., new innovation potential, decision to file for a patent), changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving). To this end, the current deliverable has been updated when the project was successfully completed so that it reflects the latest developments on data collection and management.

The rest of this document is structured as follows: Section 2 presents a series of questions along with their answers about data collection/generation. Section 3 outlines how to make DESTINI's data findable, interoperable and openly accessible. Section 4 describes the allocation of resources between the consortium partners and section 5 defines how data will be secured. Section 6 confirms that compliance with ethical principles and local and EU laws will be ensured, while Section 7 reports on any other issues. Finally, Section 8 includes an Appendix which provides a snapshot of the two datasets currently collected and used for experimentation purposes, along with a description of their main features or attributes.

2. Data Summary

In this section detailed information is reported about the contents and the collection/generation of data that was performed throughout the project. At the early stages of DESTINI, the aim was to process this information to determine the general specifications and structure of the metadata that will be generated within the DMP. At the final stage this information was validated and refined where appropriate in light of the activities that actual took place and the datasets the consortium achieved to create or study.

2.1 Q&A on Data collection/generation

1. What is the purpose of the data collection/generation and its relation to the objectives of the project?

   One of the core objectives of DESTINI's project is the investigation of a number of significant and hot topics in the field of Smart Data Processing and Systems of Deep Insight that will lead to the production of high-quality research during and after the end of the project, thus the collection/generation of datasets from the stakeholders involved play a significant role in the progress of the project.

   Smart data and systems of deep insight for locally and regionally important economic sectors will be addressed by initiating collaboration with local governmental and public bodies, so as to study problems and challenges in their line of businesses and be able in the future to setup and run various use cases of high scientific interest and great economic impact and societal value.

   Data collection/generation from DESTINI's stakeholders will constitute a greenfield for studying real-world problems by utilizing/developing new smart data processing methodologies that will be used for decision support.

   Data collection/generation initially concentrated on the following fields:

   - Transportation (e.g., marine, shipping and intelligent transport systems)
   - Agriculture – Food Industry (e.g., livestock and chicken farm management, poultry meat production & packaging)
   - Health (e.g., Personalized Medicine in the Big Data era)
   - Tourism (e.g., forecast of the availability of hotel rooms and demand)

   As the project was progressing, taking into consideration a number of factors listed below, as well as the outcome of the mid review process and in particular the suggestion

of the review committee to narrow down the areas of interest and focus on specific topics with the highest research potential and scientific value, the consortium decided to shorten the list of fields provided above so as to include only Agriculture – Food Industry and Health. The factors for this decision are as follows:

(a) Availability of the data at the time of collection activities.
(b) Willingness of owner to share it with DESTINI's consortium.
(c) Maturity of the data, that is, the format in which this data is expressed, if there is a digital form of it or not, and the effort to gather it, clean it and possibly transform it to the desired format.
(d) Value of the data, which means how appropriate and suitable is the data to facilitate experimentation with respect to the research areas of interest to DESTINI.

Based on the above, two main data collection activities were performed. The first was related also to Smart Manufacturing or Industry 4.0, and concerns the breeding of chicks in large farms. Collaboration with a major industrial unit in Cyprus has been established to monitor conditions in breeding farms, as well as other activities within the factory, such as the preparation of chicken food using an automatically controlled mill, the controlling of the climate conditions in the farms using a modern automated system, and finally the production line of packaged poultry meat, including also the activities of slaughtering, cleaning, chopping, and mixing ingredients. The second is related to healthcare and involves severe cases of patients suffering either from CA or from stroke. The associated data included MRIs, ultrasounds and a number of analyses (mostly blood) with markers and factors. A sample of the two datasets in terms of attributes and descriptors are provided in Section 8 – Appendix.

As stated in the original version of the DMP, if an opportunity would arise for experimenting with real-world apart from the aforementioned fields, and provided that the scientific field is of research and practical interest to the project, and that it falls within the general spectrum of the smart Specialization Strategy of Cyprus, the consortium would pursue it. Indeed, during the last two months of the project members of the consortium came across a rich in data and associate research challenges area, namely that of Digital Cultural Heritage. In close collaboration with one of the stakeholders of DESTINI, the Digital Cultural Heritage Research Lab and its associated ERA Chair Mnemosyne, and UNESCO Chair for Digital Cultural Heritage that were established at the Department of Electrical Engineering, Computer Engineering and Informatics at the Cyprus University of Technology (CUT), the consortium identified ample room for applying smart data processing, visualization & analytics to data pertaining the huge spectrum of Digital Cultural Heritage. To this end, the consortium started working closely with members of the DCH lab to specify structured,

unstructured, or semi-structured data of cultural heritage artifacts, ranging from buildings and monuments, to jewellery and archaeological items, for experimentation purposes using smart processing and systems of deep insight.

2. What types and formats of data will the project generate/collect?

Mainly documents (doc, docx, ppt, pptx, etc.), illustrations (png, jpeg, etc.), drawings (dwg, etc.) and numerical raw data (xml, csv, xls, .txt etc). Even though the dataset format was not initially strictly defined, the main forms of data collected contain images, videos, textual descriptions, and numerical values, while in some cases this information was enhanced by date, time, location, and the measurement results depending on the particular pilot study.

3. Will you re-use any existing data and how?

Existing data provided by DESTINI's stakeholders were and will be used in the future as input to the smart data processing methodologies that were/will be utilized within the project. More specifically, existing data were/will be used to check the accuracy of the proposed methodologies, or for solving challenges such as the efficient and fast storing and curation of data (e.g., of big data), or for supporting the decision-making process within a production line or a medical diagnosis.

For any dataset of a confidentiality nature, a non-disclosure agreement was/will be signed with the partner that owns or produces the data, while confidentiality and anonymity will be preserved in cases in which the data will be published, partly or fully, mostly in the form of papers.

In addition, any openly available dataset originated from external sources will be used for reference purposes only in order to compare the newly designed/developed approaches with a corresponding baseline. This is true for activities that will be performed past the end of the project, especially with the new topics of collaboration identified with the Early-Stage Researchers that visited Cyprus and collaborated with CUT. The main issue here is that all current datasets (Industry, Health) were classified as confidential by their owners. Therefore, only selected users will have access privileges to this data and in some cases (e.g., health records) the identity of the cases were/will be encrypted from the start, i.e., before any data is shared between researchers for experimentation.

4. What is the origin of the data?

As previously mentioned, the datasets revolved around two major areas, Industry 4.0 and health. The former concerns data from sensors and information kept regarding the preparation of chicken food through an automatically controlled mill, the controlling of the climate conditions in the breeding farms using an automated system, and finally the

production line which packages poultry meat, as well as data describing parts of the activities for slaughtering, cleaning, chopping, and mixing poultry meat ingredients. The second involves data in the form of MRIs, ultrasounds and images, as well as a number of analyses (mostly blood) with markers and factors for severe cases of patients with CA or suffering from stroke.

As previously mentioned, a new opportunity for data collection arose towards the end of the project. In close collaboration with one of the stakeholders of DESTINI, the Digital Cultural Heritage Research Lab and its associated ERA Chair Mnemosyne, and UNESCO Chair for Digital Cultural Heritage that was established at the Department of Electrical Engineering, Computer Engineering and Informatics at the Cyprus University of Technology (CUT), data pertaining the huge spectrum of Digital Cultural Heritage was identified to be of great interest and importance. To this end, the consortium started working closely with members of the DCH lab to specify structured and unstructured or semi-structured data of artifacts ranging from buildings and monuments, to jewellery and archaeological items, for experimentation purposes in applying smart data processing, visualization & analytics.

Datasets were, and are still being, collected/generated from various sources such as sensors (e.g. temperature, proximity, light, IR, humidity and distance sensors) which are already installed and used on stakeholder's production lines/premises. Sensor data will be used to make process improvements in order to increase efficiency and support decision-making. In addition, any other data that will be collected from stakeholders' desktop/web/mobile applications will be used and, if deemed useful, will also examined by the project's consortium to identify in an anonymous way how users' interact with the stakeholder's applications or users' habits and interests.

5. What is the expected size of the data?

Depending on the dataset generated/collected the expected size will vary from few kilobytes to several hundreds of MBs or few GBs. This may change in the future if more videos and images from MRIs and ultrasounds for the health-related dataset are made available to the consortium and researchers.

6. To whom might it be useful ('data utility')?

Data collected/generated is of fundamental importance for the project's consortium members, as well as for the research community dealing with smart data processing, systems of deep insight, smart data-centric services & applications. Data can be used for validation purposes, investigation and exploitation of real-world challenges and for the production of high-quality research. In addition, data can be useful for end-users as well,

such as the involved stakeholders, as it can support their day-to-day activities and their decision making.

In particular, the beneficiaries of the outcomes of experimentation and research activities will be the Paradisiotis Group in the case of the Industry 4.0 dataset, and the national health care system of Cyprus and the Cyprus Institute of Neurology and Genetics in the case of the health dataset. Additionally, beneficiaries may be regarded all other stakeholders that may utilize the approach that is currently implemented or will be implemented in the future and customize it to the needs of their line of work. This holds especially for the Industry 4.0 case. In the health-related data on the other hand, the results of simulations and processing may also benefit other patients and doctors as they will support better, more accurate and faster diagnosis. Finally, in the case of the digital cultural heritage related data, it is anticipated that a rich number of stakeholders will be benefitted by the outcomes of experimentation in the future, such as national authorities on archaeology, museums, archaeologists, monument preservators, digital cultural heritage professionals and the public.

3. FAIR Data

3.1 Making data findable, including provisions for metadata

In general, the specific typology, as well as the total number of data attributes that will exist in a dataset, cannot be precisely defined upfront to apply for every dataset. Data description depends on the typology of the collected/generated dataset and may vary from dataset to dataset. As this is a live document, every time the consortium collects/generates a dataset its data description will be updated.

In a general form the "Metadata", which is essentially a description or documentation about DESTINI's data, will consist of the following elements: "Title", "uID", "Format", "Domain", "Collection Frequency", "Data Value", "Date", "Location".

All datasets collected/generated will be well documented as this increases the prospect to be easily discovered and re-discovered by interested parties. In addition, complying with an accepted standard, this will also help in the indexing and retrieval of data.

Each metadata file will be uploaded in a standardized format. Appropriate templates will be available for download to all DESTINI's partners from the project's website under specific user privileges granted and using an authentication mechanism. In addition, metadata files

will be associated to the data descriptor, and will be directly accessible through the project's web page.

The approach followed in DESTINI is to host the data using Data Lakes (DLs). A DL enables fast storage of data produced at high frequencies and may host any kind of data, that is, structured, unstructured and semi-structured. The latter is particularly useful especially in the case of the health dataset where images and/or videos are included. The main challenge with DLs is how to organize the data so that the lake does not become a swamp. This means that we need to devise ways to retrieve easily, efficiently and effectively the information stored in a DL. To this end, DESTINI has proposed a new DL architecture that, with a new semantic enrichment mechanism and data annotation scheme, is able to tackle this challenge.

More specifically:

The architecture of our DL provides data processing (storing and retrieval) in DLs organized with pond architecture and facilitated by integrating the 5Vs Big Data characteristics and blueprint ontologies (see Figure 1). Storing and retrieval of data sources supported by visual querying providing Digital Twin characteristics.

Using the pond architecture, the DL consists of a set of data ponds each hosting / referring to a particular type of data (structured, unstructured, semi-structured). Based on the data type, each pond contains a specialized storage system and data processing. As presented on the Figure, a DL with pond architecture uses dedicated ponds to store structured, unstructured, and semi-structured data from each source. The pond architecture is particularly useful for extracting information from the DL via Visual Query.

Data Source Blueprint (DSB) is a metadata enrichment mechanism that identifies and characterizes a candidate source before it becomes a member of DLs as shown in Figure. Volume, Variety, Veracity, Value, and Velocity are the five basic characteristics of Big Data that contribute to the characterization. Basically, there are two parts to the blueprint mechanism: the stable blueprint and the dynamic blueprint. The stable blueprint records the Source Name, the Variety type and type of data, Value of the data, Velocity, and Veracity. These attributes can be assigned accordingly with standard values such as Low, Medium, High or a custom value such as frequency. Dynamic blueprint contains attributes that may change as data is processed or new data is generated.

Once this selection process is completed, the RDF created earlier becomes now part of the DL and contributes to the DL's metadata semantic enrichment which is the cornerstone for addressing the challenge of easy storing and efficient retrieval of data. The result of the query consists of the stable and dynamic blueprints selected sources, which satisfy the query parameters and thus will be added to the DL's RDF schema.

The selected data sources are then distributed to the specific DL Pond for further processing according to the corresponding attribute values. Essentially, this process and the associated characterization help to handle and manage multiple and diverse types of data sources and to contribute to the DL's metadata enrichment before and after these sources become members.

When a data source becomes part of the DL, from that point forward the metadata schema is utilized for filtering and retrieving data based on the blueprints and their metadata. The latter involves attributes such as the type of data produced by the sources, the size of the data they produce, the speed of production, the accuracy of the data, and the importance of the source data, etc. Therefore, each action for retrieving data from the DL is effectively guided by the information provided in the metadata mechanism, that is, in the blueprints.

After the completion of the selection process, the retrieval process is based on the metadata semantic enrichment – RDF schema of the DL encoded in the blueprints.

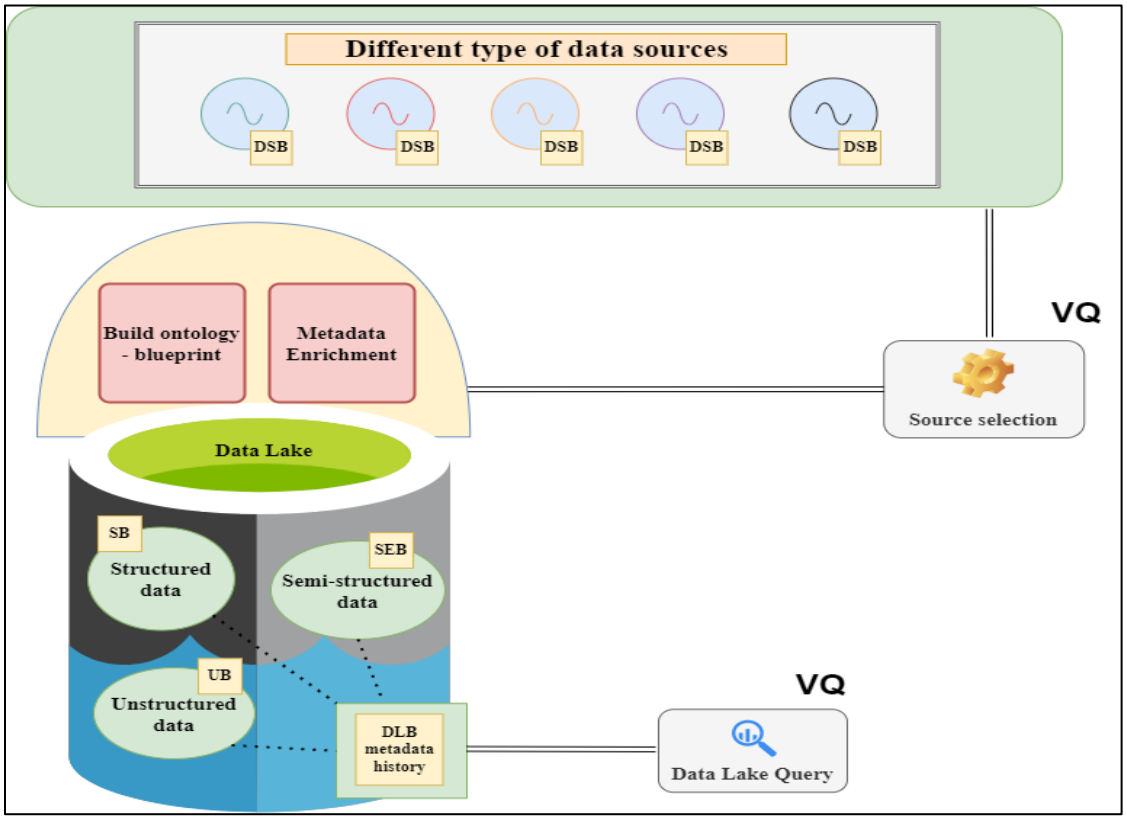

Figure 1: DESTINI's Data Lake architecture

Finally, in order to make DESTINI's datasets easy to search and locate, the subsequent identification procedure will be followed:

- Each dataset will be initially assigned with a unique ID (uID), which will be automatically generated through a Universally Unique Identifier (UUID) application. This uID will be also used in the Metadata format.
- Each dataset will be linked with a Digital Object Identifier (DOI)[1]. Using a DOI identification number ensures the uniqueness of the dataset, as well as the possibility of automatic data web retrieval.

3.2 Making data openly accessible

Only datasets related to DESTINI's scientific publications that are presented in conferences or published in well-known journals will be publicly available where feasible. Regarding other datasets collected/generated from DESTINI's stakeholders during piloting and to avoid any issues related to IP rights, the Steering Committee of DESTINI will decide on a case-by-case basis which data can be released and how.

Datasets, as well as their metadata files, will be uploaded onto the coordinator's storage located within the premises of the Cyprus University of Technology, or onto OneDrive cloud service and will be made available through the DESTINI's website.

In addition, upon approval by the Steering Committee, datasets will be also available through Open Data platforms, such as the OpenAIRE sharing web platform. This action can provide an automatic data extraction ensuring at the same time accessibility through a standard Open Data access platform.

The data acquired by DESTINI include the sectors: industry 4.0. and Healthcare. Due to GDPR reasons and NDAs, and based on the reluctance of the owners to permit opening the data to external parties, the datasets were not made freely available. We are still working on this trying to resolve it, but both stakeholders consider their data sensitive, private and business critical to allow their full sharing. A possible way-out will be the camouflage of certain features, but this is still debatable as to what benefit would bring to the research and scientific community if certain features are hidden. Nevertheless, we are still negotiating with our stakeholders, and we hope that at some point this disclosure may become feasible. Finally, if needed, we can provide an excerpt of the datasets we are currently working on, again in a closed, restricted access form, for the PO and the reviewer to view.

[1] DOI (www.doi.org)

3.3 Making data interoperable

Datasets will become available in standard/open formats which will be compliant with commercial/open software to permit any potential data exchange/transfer between researchers and institutions.

As described in sub-section 3.1, a standard vocabulary for metadata description will be used to describe datasets. If deemed necessary, a description on the most common ontologies in the area of interest will also be provided.

Finally, descriptions of the attributes present in the datasets will be made available to interested researchers and the general public. An excerpt of the two datasets collected by DESTINI thus far and its main descriptors are provided in the Appendix.

3.4 Increase data re-use (through clarifying licences)

Data will be available according to Open Licenses and upon approval by the Steering Committee.

A list of popular open licences is shown below:

1. Common Development and Distribution License
2. GNU General Public License (GPL)
3. GNU Library or "Lesser" General Public License (LGPL)
4. MIT license Mozilla Public License 2.0

As previously mentioned, in order to avoid issues related to IP rights, data will be made available for re-use purposes upon decisions of the Steering Committee. Once a dataset becomes publicly and openly available it will remain open and free.

The datasets which will be available from the beginning are the ones reported in scientific publications originated from members of the consortium, which are essentially made for reuse. Finally, as regards to data quality, this is assured by each partner, who bears full responsibility of the collected/ generated data.

4. Allocation of resources

The costs for making data FAIR include the costs of the web/cloud storage, as well as the costs of personnel involved in collecting, generating and managing data. The Cyprus University of Technology (CUT), as the coordinating institution, will be the main partner responsible for collecting/generating data from involved stakeholders in Cyprus. The rest partners, UNIROMA and ERISS/JADS, will also assist in the data collection process by either gathering data from stakeholders in Italy or Netherlands who are interested in supporting the DESTINI project, or contributing to validation and exploitation of data collected in Cyprus.

Other activities where an allocation of personnel is necessary include:

- Set up of the data storage on the server located within the Cyprus University of Technology
- Set up of a special section within the project's website where authorized users will be able to add/edit/delete data
- Preparation of templates for the data descriptors, as well as for the format of the metadata.
- DOI registration request
- Upload of data on Open Data platforms such as OpenAIRE

5. Data Security

Data is currently, and will be in the future, stored on servers of the Cyprus University of Technology which is part of its cloud storage infrastructure. Where appropriate, data may be shared with interested researchers and other stakeholders through the project website.

DESTINI differentiates between three kinds of research data:

a) Open: This refers to data which are freely available online. This mostly refers to deliverables and research publications (preliminary camera-ready versions).
b) Controlled: Involves data with restricted access based on any ethical, legal and/or commercial reasons prohibiting their open release. This is partly the form of the data released between closed groups of researchers within the consortium.
c) Closed: Entails data which are not open to any sort of experimentation due to their nature. This kind of data was avoided in DESTINI, especially with health records.

It should be noted that only users with authorized access (username + password) will have the rights to view/edit/add a dataset as per category (b) above. Especially with the datasets collected thus far (industrial and health data) only a limited group of researchers within the consortium have access. In addition, any confidential or private information has been either encrypted or removed completely.

6. Ethical Aspects

The Consortium confirms that compliance with ethical principles and applicable international, EU and national law in the implementation of research activities will be ensured.

7. Other Issues

There are no other issues to be documented.

APPENDIX

1. INDUSTRY4.0 RELATED DATASET – Column Descriptors

An excerpt of the dataset of Paradisiotis Group is provided here, which was acquired through a dedicated system called Agrologic that controls climate conditions within the chicken farms.

The columns are related to measurements, daily or hourly, of the following attributes:

i. Grow day
ii. Battery temperature
iii. Humidity - current day max.
iv. Humidity - current day min.
v. Required temperature
vi. Temp. in - current day max.
vii. Temp. in - current day min.
viii. Temp. out - current day max.
ix. Temp. out - current day min.
x. Heater 1 daily on time (hh:mm)
xi. Heater 2 daily on time (hh:mm)
xii. Heater 3 daily on time (hh:mm)
xiii. Reset time
xiv. Number of birds

INDUSTRY4.0-RELATED DATASET – Data Snapshots

Grow day	Battery temperature	Humidity - current day max.	Humidity - current day min.	Required temperature	Temp. in - current day max.	Temp. in - current day min.	Temp. out - current d
1	35	39	30	33,5	34,4	33	
2	35	46	32	32,6	34,3	31,1	
3	36	43	28	32,3	33,2	30,3	
4	36	34	25	31,9	32,9	29,4	
5	35	36	28	31,5	32,3	29,5	
6	36	35	21	31,1	32	29,2	
7	36	42	17	30,7	31,7	29	
8	35	40	18	30,3	31,6	28,8	
9	35	35	17	29,9	31,6	28,5	
10	35	35	17	29,5	31,5	28,2	
11	36	40	20	29,1	31,3	28,3	
12	36	41	29	28,7	30	28,2	
13	36	51	38	28,3	29,2	27,5	
14	36	50	38	27,9	28,8	27,6	
15	36	54	36	27,5	28,8	27,2	
16	36	50	34	27,2	28,1	26,8	
17	36	56	39	26,8	27,7	26,4	
18	36	53	40	26,4	27,2	25,9	
19	36	52	40	26,1	26,7	25,6	
20	35	52	41	25,7	27	25,2	
21	35	53	40	25,5	26,8	24,9	
22	35	54	40	25,4	27,2	24,9	

Grow day 1 Hour(24)	Required temperature	Sensor 3	Sensor 4	Outside temp.	Current average temp	Humidity (%)	Dimmer output (%)	Current CO2 (PPM)	CO2 hour maximum	CO2 hour mi
0	0	0	0	0	0	0	0	0	0	
1	32,5	32,5	32,9	22,3	32,8	43	50	549	549	
2	33,5	33,7	33,6	24,7	33,9	42	50	588	627	
3	33,5	34,4	33,9	27,6	34,4	42	50	627	666	
4	33,5	35	34,7	30,2	34,9	40	50	509	627	
5	33,5	35	35,1	31,8	35	41	50	431	549	
6	33,5	35	35,7	33,5	35,2	44	50	470	470	
7	33,5	34,4	35,1	32,1	34,6	46	50	470	470	
8	33,5	34,4	35	32,3	34,6	46	50	470	470	
9	33,5	34,2	34,9	31,8	34,6	47	50	509	549	
10	33,5	34,3	34,9	29,6	34,6	48	50	627	666	
11	33,5	34,1	34,5	30,9	34,4	43	50	705	745	
12	33,5	33,3	34	26,7	33,7	42	50	705	745	
13	33,5	33,1	34,1	25,5	33,7	40	50	745	745	
14	33,5	33,2	34,1	24,3	33,7	38	50	705	745	
15	33,5	32,9	34	23,7	33,6	37	0	666	745	
16	33,5	32,9	34	22,7	33,6	40	50	745	745	
17	33,5	32,7	33,9	22,5	33,4	40	50	745	784	
18	33,5	32,7	34	22	33,5	40	50	705	784	
19	33,5	32,5	33,9	21,4	33,4	40	50	745	784	
20	33,5	32,6	34	20,9	33,5	40	50	745	784	
21	33,5	32,6	33,9	20,9	33,4	40	50	784	823	

2. HEALTH-RELATED DATASET – Column Descriptors

Ar_cl_lm: Black area close to lumen, this is a feature described by the team of colleagues. It is included in:

- Guidelines published by the European Society for Vascular Surgery (ESVC): "Editor's Choice - Management of Atherosclerotic Carotid and Vertebral Artery Disease: 2017 Clinical Practice Guidelines of the European Society for Vascular Surgery (ESVS)" , published in Eur J Vasc Endovasc Surg (2018) 55, 3-81.
  Discrete white areas: shows areas displayed as white(calcified) above a certain size. , this is a feature described by our team. It is included in (attached is the original paper):
- Guidelines published by the European Society for Vascular Surgery (ESVC): "Editor's Choice - Management of Atherosclerotic Carotid and Vertebral Artery Disease: 2017 Clinical Practice Guidelines of the European Society for Vascular Surgery (ESVS)" , published in Eur J Vasc Endovasc Surg (2018) 55, 3-81.

Column F (Kurtosis) up to column AL (SGLDM Angular Second Moment) are described in the "Carotid Plaque Texture Analysis Research Software for Ultrasonic Arterial Wall and Atherosclerotic Plaques Measurements" manual, Appendix B (available upon request)

Column AM : this is the size of the plaque in square millimeters

Column AN: shows the stenosis of the artery using the ECST method (standardized procedure followed by ultrasonographers)

Column AO : this is the percentage of black area in relation to the plaque in the first component of the plaque

Column AP : this is the percentage of black area in relation to the plaque for cases we have two components of plaques

Column AQ: The total size pf black area in square milimeters

HEALTH-RELATED DATASET – Data Snapshots

Img_name	Ar_cl_Im	Pat_Stat	Discr_white_areas	Pl_Type	Kurt	SGLD_ASM	SGLD_CON	SGLD_COR	SGLD_VAR	SGLD_HOM	SGLD_SAV	SGLD_SVA	SGLD_SEN	SGLD_ENT	SGLD_DVA	SGLD_DEN	SGLDInM1	SGLDMInM2	MEAN	VARIAN	MEDIAN	MODE	
1	Yes	A	Yes	2	1,7765957	0,0414636	31,7689812	0,9875462	1276,451618	0,417815	50,645167	5074,037491	4,5245587	6,1576687	20,0194532	2,3124701	-0,4267889	0,981409	24,3651894	1278,213812	10,8793103	0	2,
2	No	A	Yes	2	2,0463171	0,0145571	22,4275806	0,9738799	429,3240607	0,3586035	37,293311	1694,868662	4,4825756	6,2877202	11,778868	2,2791425	-0,3571805	0,9650573	17,6670262	428,7514837	10,7821351	0	2,
4	Yes	S	Yes	2	1,8403834	0,0341686	27,339156	0,9740736	527,2346718	0,3755044	38,1310246	2081,599531	4,3556532	6,1196865	15,366228	2,3386798	-0,3549066	0,9621635	18,0738942	526,824891	10,6272912	0	2,
10	Yes	A	Yes	2	2,4061514	0,0066964	32,2050298	0,9758691	667,705246	0,3102382	59,7690065	2638,615954	4,9728312	7,0076316	17,4468753	2,4117563	-0,3737718	0,9785367	29,0407505	670,2698575	23,5144788	0	1,
11	No	A	Yes	3	1,6970744	0,0008755	49,2573528	0,9796001	1207,876046	0,205377	97,1631038	4782,246832	5,4681203	7,9012465	21,2228464	2,6752458	-0,3480139	0,9807002	47,5035151	1209,046616	39,5652819	0	0,
12	No	A	Yes	3	1,8888542	0,0034147	111,8145332	0,9473055	1060,017149	0,2143506	87,495952	4128,254063	5,2634762	7,6473528	49,9705933	2,9856568	-0,3249978	0,9709568	42,4373188	1053,184477	37,4090909	0	1,
13	Yes	A	No	1	6,5637074	0,4833233	16,0150587	0,8744866	64,096244	0,7610796	9,0053308	240,3699174	1,8488965	2,418863	12,933209	1,2981043	-0,3360779	0,7834966	3,5847157	66,0367313	0	0	2,
14	No	A	Yes	2	2,5652787	0,0264927	88,2593075	0,9039004	459,3572252	0,3298499	53,0717195	1749,169593	4,5578761	6,5903904	43,9622529	2,7583764	-0,2796603	0,9323129	25,5223765	459,7640236	23,2327586	0	0,
15	Yes	T	Yes	2	1,6989705	0,0021151	24,7271944	0,9883295	1059,045052	0,3074482	72,8454816	4211,453014	5,2428776	7,2366385	12,4645503	2,3282129	-0,4133744	0,9878007	35,3601374	1054,044206	25,1242884	0	1,
16	No	A	No	3	1,7473882	0,0010491	39,8893827	0,9877602	1634,396115	0,2550116	94,1847509	6497,695078	5,4467279	7,4430823	18,6863817	2,5157587	-0,4429697	0,99197	46,1469662	1648,359673	32,9302326	17	1,
17	Yes	S	Yes	2	2,4669731	0,0590204	63,4610501	0,9409815	538,3406011	0,385519	46,0440676	2089,901355	4,3235352	6,1708578	33,9328672	2,5704818	-0,3123263	0,9410443	22,0507072	541,3194208	15,0852535	0	1,
18	Yes	S	Yes	3	1,8019712	0,003187	37,3317594	0,9933995	2827,777047	0,3147632	122,3720828	11273,77643	5,6933931	7,7908595	20,8264752	2,4590272	-0,4476174	0,9938622	60,1587099	2827,152845	46,820602	0	0,
19	Yes	S	Yes	3	1,7999634	0,0160685	53,9443555	0,9924774	3583,584151	0,378518	108,5013915	14280,39225	5,3492467	7,1926586	35,7755151	2,4602337	-0,4731224	0,9936255	52,9782571	3578,584629	28,7260638	0	1,
20	No	T	Yes	3	1,7356407	0,0012926	481,4264521	0,8198346	1321,13967	0,1748839	92,552982	4803,132227	5,4090088	8,0061842	386,1251132	3,1104355	-0,3056993	0,9703604	44,440446	1165,1049	38,9473684	0	1
21	Yes	S	Yes	3	1,7999634	0,0134998	71,8856122	0,9868497	2733,739057	0,3474491	93,4640719	10863,07062	5,2922609	7,3374968	47,2972961	2,5856304	-0,4187865	0,9883623	45,66043	2732,967342	26,4798206	0	1,
23	Yes	T	Yes	2	1,8089631	0,131321	35,617152	0,9878195	1461,544761	0,5326387	48,8046871	5810,561891	3,9743223	5,3456587	25,9397146	2,0981713	-0,4545387	0,9766834	23,3693283	1457,491701	4,3588537	0	2,
25	No	A	Yes	4	1,7640356	0,001253	36,9896396	0,9702825	622,4814407	0,278104	109,1873272	2452,936123	5,1406838	7,2312834	20,2548902	2,3917511	-0,377699	0,9800116	53,74033111	625,5904142	48,9782609	41	1,
33	Yes	S	Yes	2	1,829274	0,0035104	38,4835728	0,9807954	1001,161593	0,3194626	69,1705693	3966,1628	5,1548061	7,1618492	23,3856774	2,4076074	-0,400678	0,985018	33,540177	994,0335751	23,8007663	0	1,
47	No	S	Yes	3	1,8041644	0,0008324	105,609719	0,9712868	1838,266837	0,2581061	110,4683433	7247,457628	5,6065724	7,969496	66,0426519	2,7555473	-0,3806899	0,9854995	54,1801539	1830,450451	42,4145907	16	1,
51	Yes	S	Yes	3	1,7260582	0,0053509	92,3042158	0,9719249	1652,877437	0,3142099	106,0581479	6519,205532	5,4291092	7,6069283	59,3474762	2,6705743	-0,4046314	0,9869501	52,4352999	1685,257932	50,079235	0	0,
54	Yes	S	Yes	1	1,7954544	0,1305367	13,5308592	0,9844713	438,9934587	0,6777382	23,3279933	1742,442976	3,4536926	4,2984093	11,4523716	1,5029856	-0,5497899	0,9799515	10,8029607	457,1232702	2,7284916	0	5,
90	Yes	T	Yes	3	1,7470376	0,0015378	41,8411532	0,989216	1940,221337	0,2856005	118,0026895	7719,044193	5,6344984	7,7351591	22,7525484	2,5267753	-0,4383632	0,99305	57,7775472	1940,511081	52,3526316	0	0,
100	No	A	Yes	3	1,7963095	0,0005088	212,1273767	0,9658342	3099,565563	0,1551326	140,0612053	12186,13488	5,7891254	8,1124983	121,23937	3,1876757	-0,4139076	0,9922711	68,6556623	3060,646338	52,0405405	21	1,
101	No	A	No	4	1,7356259	0,0008047	181,6366934	0,9647629	2567,076979	0,1696865	155,146641	10086,67122	5,7184654	7,6542851	92,5683047	3,1647458	-0,4752707	0,9955245	75,2285714	2529,261236	60,4210526	33	0,
102	No	A	Yes	3	2,3854041	0,0011711	87,2800317	0,9497628	871,1404484	0,2152399	81,6887561	3397,281762	5,1972979	7,3103873	46,6725402	2,8263039	-0,3836067	0,9831295	39,6438596	878,8228717	31,328125	22	1,
103	No	A	Yes	3	1,831984	0,0007809	90,5926639	0,9701841	1519,111081	0,184827	126,9326986	5985,851659	5,5296671	7,5882949	42,9798821	2,9064142	-0,4471252	0,9934446	61,8869286	1521,320816	57,2096774	21	0,
104	No	A	Yes	3	1,8259587	0,0018711	339,196651	0,9517351	3526,903357	0,168063	144,8127607	13768,41678	5,7216654	7,5120831	216,1987336	3,2777876	-0,5168181	0,997244	70,5607196	3581,115878	58,4722222	0	0,
105	Yes	A	No	1	4,998783	0,0788812	16,6740319	0,9273521	114,7798503	0,4964894	16,2148176	442,4453694	3,378276	4,5681506	10,3431338	2,0277502	-0,3012511	0,8917444	7,1644384	116,3513156	3,0932056	0	2,
106	Yes	A	No	2	2,1203936	0,0646598	17,0774342	0,9802541	436,8615605	0,5538567	25,4111483	1730,368808	3,6999118	4,7784234	12,0640386	1,8787135	-0,4446871	0,9644974	11,836541	454,9132583	3,1634241	0	3,
107	No	A	No	3	1,9396585	0,0012756	59,1025396	0,9706085	1003,334438	0,2158391	97,097925	3954,235212	5,2695573	7,2940274	27,1131479	2,7130346	-0,4116116	0,9877454	47,3392792	986,3166333	39,5446429	18	1,
108	No	A	Yes	2	2,1160254	0,074832	56,4518727	0,9526109	594,6885416	0,4474141	39,966639	2322,302294	4,1120307	5,6091186	36,9419376	2,4026555	-0,3807509	0,9607433	18,9962619	589,1048669	10,5837563	0	2,
109	No	A	Yes	3	1,7869649	0,0005362	200,8861639	0,951121	2059,715609	0,1664167	113,282958	8037,976273	5,5936325	8,0792087	109,8016906	3,1678669	-0,3597568	0,9836012	56,1226803	2082,295192	41,6161616	16	1,
110	No	A	No	4	1,7647635	0,0004722	348,0248332	0,9100665	1948,122997	0,1208208	244,7002205	7444,467155	5,6954607	7,8325293	168,6060373	3,4120824	-0,4580759	0,9949353	120,142315	2008,502517	114,875	107	0,
111	No	A	Yes	3	1,7015346	0,0005059	567,1938237	0,9086035	3090,86947	0,1028189	178,0238785	11796,28405	5,7872985	7,7362678	299,7948517	3,6143614	-0,504449	0,9972231	86,2247874	3047,439143	78,4285714	30	0,
112	No	A	No	3	1,7316332	0,0008928	196,2045119	0,9583426	2360,005117	0,1699727	122,4119309	9243,815956	5,4375208	7,3862932	102,6925565	3,1409689	-0,4712264	0,99438	59,5316387	2364,965062	39,15	25	1,
113	Yes	A	Yes	3	2,1808772	0,000987	226,6423438	0,9048393	1189,508339	0,16462	112,6096754	4531,391011	5,347195	7,1926704	114,4357491	3,2204689	-0,4835158	0,9947008	54,0246262	1181,221909	50,625	37	0
114	No	A	Yes	3	1,8066377	0,0004441	355,2311057	0,8945664	1680,268094	0,1298032	136,3265487	6365,841272	5,6284826	7,9500192	184,6879258	3,4642291	-0,4080626	0,9911572	66,3490034	1665,076391	61,5609756	60	0,
115	No	A	Yes	2	2,1018461	0,0019779	86,0901941	0,9322171	638,615462	0,2415509	58,8361923	2468,371654	4,9837182	7,1789881	48,063496	2,789884	-0,3345954	0,9680682	28,7046736	667,4305128	22,5474138	0	1,
116	No	A	Yes	3	2,3156271	0,0011184	154,6233584	0,8952563	738,1007898	0,1876313	74,2271701	2797,779801	5,1391048	7,4553636	82,0495318	3,0662662	-0,3348202	0,9716796	36,3589514	736,5464592	30,5641026	13	1,
117	No	A	Yes	3	1,7999634	0,0053204	401,8121974	0,9479176	3908,258167	0,1916213	127,3883927	15231,22047	5,6582238	7,9036892	244,6254042	3,424317	-0,4133346	0,9912488	64,7847156	4193,661787	42,1122449	0	1,

MODE	SKEW	ENERG	ENTR	GLDM_HOM	GLDM_CON	GLDM_ENE	GLDM_ENT	GLDM_MEA	fps_frad	fps_fang	runl_SRE	runl_LRE	runl_GLD	runl_rld	Runl_RP	Area_mm2	Perc_sten	Perc_black_area_1_comp	Perc_black_area_all_plaque	Black_area_m	
0	2,6075287	0,0638073	3,9173802	0,4180631	31,7041146	0,1503042	2,3244149	3,3625679	2278,360986	2281,151949	0,8966984	2,407871	297,6931569	12775,02717	16,1555176	59,12	85	22,25	15,48	9,15	
0	2,0984268	0,0394598	3,8283325	0,3587181	22,4063701	0,1314689	2,3010084	3,1962007	2212,784469	1928,047299	0,9058589	2,1092716	405,1986389	13262,53756	16,4157715	53,31	75	5,68	5,68	3,03	
0	2,4036163	0,0627087	3,7184835	0,3756495	27,2993995	0,1322571	2,3528274	3,4038956	2385,107266	2025,720049	0,9045685	2,2962074	490,5122576	17051,29457	21,1706543	73,08	75	11,9	11,9	8,7	
0	1,1672475	0,0203455	4,3132539	0,3106774	32,098409	0,113739	2,4344696	3,7564361	2764,674856	2490,774761	0,9223608	1,7041839	376,6964666	22048,27287	26,2185059	80,61	75	13,57	7,54	6,08	
0	0,9387076	0,0099328	4,7825133	0,2055872	49,1580519	0,0832222	2,7032334	5,1895764	4058,288459	3193,507821	0,9464648	1,3581031	241,7259509	22163,10436	24,8393555	69,7	80	3,59	1,28	0,89	
0	1,2022942	0,0159131	4,5583935	0,2147931	111,5571982	0,0634699	3,0419813	7,5857973	3242,070717	2394,909309	0,9384256	1,5967208	114,8258668	8282,975291	9,3967285	27,6	80	6,2	6,2	1,71	
0	2,8676495	0,5477552	1,4720515	0,7612404	15,9903367	0,5323428	1,3100822	1,6923574	1099,474733	1090,477107	0,8541882	4,4089855	382,8871272	5797,623484	8,0356445	78,22	80	52,43	52,43	41,01	
0	0,6995438	0,0523869	3,8293177	0,3307332	87,8934116	0,098434	2,80279	6,4038229	2022,333742	2095,953928	0,9076347	2,2055589	326,1662182	13299,57809	16,2390137	55,08	83	14,18	3,46	1,91	
0	1,2659265	0,0137972	4,5592276	0,3075569	24,7044599	0,1254016	2,3444902	3,4450101	3386,087042	3485,597806	0,914518	1,7047312	447,7650003	29225,87665	35,4821777	106,27	50	7,18	7,18	7,63	
17	1,1649322	0,0106346	4,781742	0,255774	39,726017	0,1016734	2,5669549	4,4266753	2397,299846	2062,798983	0,9249118	1,4553537	43,434921	3521,272677	4,1479492	11,91	85	8,44	8,44	1,01	
0	1,1403938	0,0890802	3,6581122	0,3860874	63,2739453	0,129603	2,6143296	5,2182653	2385,879653	1881,719208	0,9102302	2,3531064	557,6193975	24445,97689	29,6345215	110,29	94	33,43	13,19	14,55	
0	0,9120563	0,0103185	5,0166832	0,3150265	37,2540133	0,1135969	2,4849732	3,9585423	5989,020681	4663,255416	0,913615	1,8768162	325,0579931	37476,36914	45,4758301	140,77	90	14,66	7,29	10,26	
0	1,2165385	0,0282762	4,704461	0,378871	53,7889027	0,1290735	2,4881447	4,0998224	5058,620677	4266,036773	0,8970558	2,3442368	188,0482226	16652,31203	21,0141602	71,95	90	22,22	22,22	15,99	
0	1,105085	0,0119755	4,699272	0,1773848	175,8426374	0,0646647	3,0502613	7,7034456	3335,679202	2930,849622	0,953376	1,3403493	131,9318926	11264,6768	12,3984375	34,53	94	7,86	7,86	2,71	
0	1,4450437	0,0264863	4,6371609	0,3477658	71,6888585	0,1143537	2,6173995	4,750967	6996,602826	7023,574038	0,9078509	2,0164231	416,4531418	33130,98037	40,7424316	133,56	97	6,23	4,84	6,47	
0	2,2571767	0,1564868	3,4597734	0,5329767	35,5054211	0,2282103	2,1181944	2,9874854	2845,462921	2584,069772	0,8999015	2,4680037	343,733235	16710,57073	20,9001465	95,45	90	23,95	17,75	16,94	
41	1,4379704	0,0140928	4,4612113	0,2786718	36,876597	0,1158423	2,4444323	3,9237456	3706,573634	2512,3585	0,9181992	1,6249025	168,2831082	9994,128207	11,9448242	35,04	83	1,08	1,08	0,38	
0	1,2866144	0,0162631	4,4777051	0,319764	38,4155484	0,1230042	2,4292308	3,7995832	3359,302068	2184,667847	0,9084871	1,8214618	218,1281637	13288,45642	16,3630371	50	80	13,85	13,85	6,92	
16	1,3879462	0,0091254	4,9202902	0,2585976	105,2896997	0,0919374	2,8108969	5,9415462	4537,773343	3359,51362	0,9266346	1,617457	218,4087205	21096,58495	24,6940918	72,12	85	12,57	5,99	4,32	
0	0,7218451	0,0140806	4,7792554	0,3148885	91,9128985	0,1026896	2,7268729	5,3887522	4289,900268	3624,367635	0,9166415	1,9146705	253,488744	24263,77327	29,0527344	91,19	99	37,41	24,03	21,91	
0	5,1819294	0,1550655	2,9742736	0,6779457	13,4945867	0,355158	1,5105834	1,3989211	1438,035609	1314,744733	0,8230216	3,9647278	537,6424862	11762,47705	17,8803711	101,25	85	46,3	30,87	31,26	
0	0,6181636	0,0088214	4,9492101	0,2857458	41,7883374	0,1052311	2,5497432	4,2731299	4779,560557	4944,274682	0,9145795	1,6847193	101,3896125	10910,49621	13,2502441	39,75	90	17,92	17,92	7,12	
21	1,1789381	0,007709	5,1110829	0,1559326	211,0667361	0,0533793	3,2402583	9,1692628	2029,521686	1687,980279	0,9541788	1,2871248	34,3944279	4153,100017	4,5473633	12,51	90	7,67	3,72	0,47	
33	0,8789632	0,008713	5,0116185	0,1711853	180,1646316	0,0528352	3,229323	9,0309474	4366,584333	1954,459934	0,9388555	1,4010045	18,2578791	1943,102395	2,2043457	6,21	65	4,63	4,63	0,29	
22	1,3186003	0,014277	4,5229805	0,2169012	86,5457844	0,0771367	2,8719764	6,1208308	1077,202348	825,2579495	0,9313408	1,4002826	42,1381364	2605,799669	3,0124512	8,55	60	17,59	8,63	0,74	
21	0,8262904	0,0089054	4,8875455	0,186033	89,837407	0,0673529	2,950674	6,6763018	1507,587019	1109,9007	0,9439645	1,3287623	19,4053486	1987,812795	2,2351074	6,24	75	8,53	5,81	0,36	
0	0,9178021	0,0095222	5,0612979	0,1688446	336,6217488	0,0481954	3,3451402	10,5702523	2599,266722	2378,313722	0,9495763	1,4148243	13,413753	1615,45941	1,791748	5,06	90	1,09	1,09	0,06	
0	2,2270757	0,1583449	2,6983387	0,4967751	16,6403679	0,2012717	2,0425147	2,4650963	1319,437529	1277,069035	0,8135129	3,7590963	386,4933044	3655,606867	5,7280273	23,75	80	35,96	35,96	8,54	
0	3,3670757	0,1206338	3,0754856	0,554317	17,0303694	0,2396644	1,9026381	2,1392657	1408,70876	1285,114911	0,7906666	4,5512069	311,3200033	3911,187564	6,4191895	28,45	75	79,08	79,08	22,5	
18	1,2138574	0,013333	4,5947489	0,2174206	58,6231059	0,0796684	2,7594149	5,461828	1183,878753	938,2260203	0,933953	1,435267	30,5509061	2094,436604	2,4133301	6,87	97	4,22	1,89	0,13	
0	2,2945014	0,1065379	3,4712241	0,4479816	56,2661543	0,1674407	2,4305345	4,2467501	1565,538472	1464,457179	0,8838197	2,4935481	127,2068337	4459,762602	5,7634277	23,41	80	6,21	6,21	1,45	
16	1,2406144	0,0092099	4,934606	0,1672555	199,7018186	0,0559188	3,2353052	9,0439826	2441,831931	2011,024693	0,9506259	1,2883571	58,7420706	5787,482264	6,3964844	17,65	80	3,91	3,22	0,57	
107	0,3742361	0,0070167	5,0918058	0,1229358	342,8597095	0,0415534	3,6299184	12,3829001	2072,649431	1802,959239	0,968052	1,158335	14,0159939	1859,56164	1,9672852	5,27	90	0,31	0,19	0,01	
30	0,3936483	0,0063647	5,1666306	0,1038474	561,9393312	0,0344214	3,7241308	15,2906051	2410,353119	2169,935019	0,9714443	1,1559672	9,9892312	1469,072894	1,541748	4,12	85	0,61	0,61	0,03	
25	1,2577143	0,0111551	4,818956	0,1715767	193,7624655	0,05922	3,2216614	9,0434656	1482,983145	904,0870326	0,9542708	1,2501888	18,5022331	1538,54173	1,6867676	4,62	80	12,52	5,52	0,26	
37	0,476841	0,0099608	4,7354139	0,1654008	225,3102057	0,051409	3,320805	9,9534614	1283,245684	746,1732279	0,9494374	1,2827703	10,3216694	927,9831633	1,0292969	2,84	80	7,39	7,39	0,21	
60	0,6175452	0,0078182	4,9878313	0,1310671	350,6536688	0,0423847	3,5289133	12,3451283	1847,277331	1409,263929	0,9637047	1,196101	19,5666857	2295,413522	2,4604492	6,65	80	11,33	7,52	0,5	
0	1,6202233	0,0171306	4,3230006	0,2420674	85,812235	0,082749	2,8393494	5,9001072	1537,782414	1130,686225	0,9224097	1,5609117	75,8085374	3985,250083	4,723877	13,75	85	11,86	11,86	1,63	
13	1,1002735	0,0137511	4,4819678	0,1885135	153,7679205	0,0623689	3,1319503	8,0799291	1460,344558	1391,233797	0,9447116	1,3761843	40,2262992	2673,313951	2,9941406	8,39	80	4,26	4,26	0,36	
0	1,1326203	0,0144191	5,0091319	0,1926341	398,0954075	0,0499113	3,4678783	11,886591	1907,467042	1482,215421	0,9448308	1,6443362	43,3912561	4994,026166	5,5876465	16,41	90	2,47	1,81	0,3	