



## **INSANE - The effects of urbanization on INsect diverSity And humaN-nature intERactions**

### **Data Management Plan**

|                                 |   |
|---------------------------------|---|
| <b>Project Name</b>             | The effects of urbanization on insect diversity and human-nature interactions |
| <b>Project Acronym (number)</b> | INSANE (101033024)  |
| <b>Funding</b>                  | European Commission - MSCA-IF-EF-CAR (2020)                                   |
| <b>Coordinator</b>              | <b>Olivia Sanllorente</b>   |
| <b>Beneficiary</b>              | <b>University of Granada, Spain</b>   |
| <b>Deliverable Number</b>       | 1   |
| <b>Deliverable Title</b>        | Data Management Plan  |
| <b>Due Date</b>                 | 28/02/2022  |
| <b>Issue Date</b>               | 11/02/2022  |
| <b>Status</b>                   | Submitted   |
| <b>Dissemination level</b>      | Public  |
| <b>Contact</b>                  | <a href="mailto:oli@ugr.es">oli@ugr.es</a>                                    |

| <b>Version</b> | <b>Modifications</b>      | <b>Date</b> | <b>Author</b>                      |
|----------------|---------------------------|-------------|------------------------------------|
| 1.0            | Initial Document Creation | 11/02/2022  | <a href="#">Olivia Sanllorente</a> |

## Contents

|   |    |
|---|----|
| 1. Data Summary.....                                  | 3  |
| 1.1. Introduction and purpose .....                   | 3  |
| 1.2. Relevance to project targets.....                | 3  |
| 1.3. Type, origin, format, and quantity of data ..... | 4  |
| 1.4. Data value .....                                 | 7  |
| 2. FAIR Data .....                                    | 8  |
| 2.1. Making data findable .....                       | 8  |
| 2.2. Making data openly accessible .....              | 9  |
| 2.3. Making data interoperable .....                  | 10 |
| 2.4. Increase data re-use.....                        | 11 |
| 3. Allocation of Resources .....                      | 11 |
| 4. Data Security .....                                | 12 |
| 5. Ethical Aspects.....                               | 12 |
| 6. References .....                                   | 12 |
| 7. List of Acronyms.....                              | 12 |
| 8. Annex A. Work Packages .....                       | 13 |

## 1. Data Summary

### 1.1. Introduction and purpose

This report describes the initial **Data Management Plan** (DMP) for the project **INSANE (101033024)**, which is funded by the European Commission through the Marie Skłodowska-Curie Actions (MSCA) as part of the Individual European Career Restart Fellowship Programme (IF-EF-CAR). The coordinator of this DMP, who is also the experienced researcher associated to the funded fellowship, is [Dr. Olivia Sanllorente](#) from the [department of Zoology](#) at the [University of Granada](#) (UGR), Spain (the beneficiary institution).

The purpose of this DMP is to provide a detailed description of the procedures and protocols for the management of the datasets generated during the lifetime of the project. This DMP will describe the main data management principles in terms of data standards and metadata, sharing, archiving, preservation, and security.

This is an alive document that will be updated at regular intervals during the lifetime of the project and will be allocated in the institutional repository of the [UGR](#), [DIGIBUG](#), under the name of **INSANE\_01\_DMP\_V1.0\_WP5.pdf** (see [section 2.1](#) for naming conventions).

### 1.2. Relevance to project targets

INSANE will generate several datasets of different types both quantitative and qualitative. The data management will serve to support the project scientific objectives and spread the project results. This includes the management of three main data categories:

**1- Research objectives.** The datasets associated to this category will allow any potential user to replicate the main scientific results of the project. This includes data from field work, surveys, DNA barcoding and computer simulations as well as codes to produce and analyze data.

**2- Dissemination activities for expert audiences.** The datasets associated to this category will give access to any potential user to those documents summarizing the main scientific results of the project. This includes preprints, technical reports (*e.g.*, protocols), and conference presentations.

**3- Communication activities for non-expert audiences.** The datasets associated to this category will give access to any potential user to those documents dedicated to educational purposes. This includes presentations, materials and publications used in events for different non-expert audiences and teaching material for undergraduate and graduate students.

### 1.3. Type, origin, format, and quantity of data

INSANE aims to provide novel and useful information on how to reconcile urban development and nature conservation, particularly investigating the impact of urbanization on insect diversity and human-nature interactions. To this purpose, INSANE proposes two complementary approaches. On one hand, the project INSANE will generate data coming from two different experimental approaches: field work to obtain insect diversity (taxonomical data) and citizens' surveys to obtain their perception of wellbeing and extinction of experience (sociological data). On the other hand, INSANE will also generate data coming from Molecular Genetics (metabarcoding) and Machine Learning, whose codes will be created from scratch and will be open to any potential user. Apart from the research objectives, INSANE will generate data to disseminate its results between different expert audiences and data for educational purposes. A brief description of the main objectives (both scientific and pedagogical) can be found in the work package table presented in [Annex A](#).

INSANE will generate data with different formats that will be accessible using free software (see table in [section 2.2](#)). The different formats are presented in the table below.

| Type of Data                         | Description  | Format  |
|--------------------------------------|--|---|
| <b>Compressed data</b>               | Apart from saving data storage, compression will be used for packaging files with similar and/or complementary content | <b>TAR.GZ</b><br>(other formats inside)                                 |
| <b>Raw numerical data</b>            | Experimental and computational   | <b>dat/csv</b>  |
| <b>Images</b>                        | Scientific and educational purposes  | <b>JPEG, PNG, TIFF, PDF</b>   |
| <b>Codes (programming languages)</b> | Used for both running simulations and analyzing data   | <b>Python</b> (codes), <b>dat/csv</b><br>(parameter files and readme's) |
| <b>Reports</b>                       | Scientific, technical, and educational purposes  | <b>PDF</b>  |
| <b>Paper preprints</b>               | <a href="#">Green open access according to the H2020 guidelines</a>  | <b>PDF</b>  |
| <b>Graphical representations</b>     | Used for graphical analysis and Molecular Visualization  | <b>XMGRACE, VMD</b>   |
| <b>Presentations</b>                 | Scientific and educational purposes  | <b>PDF and Keynote</b> (talks and posters), <b>mp4</b><br>(video)       |

INSANE will create data from **six distinct origins**: **fieldwork** (biodiversity, environmental and sociological), **DNA barcoding**, **computational**, **reports** (technical and pedagogical), **paper preprints**, and **presentations** (technical and pedagogical). In the following four tables we summarize in a tentative scheme the different types of datasets that will be generated by INSANE linked to the five project work packages presented in [Annex A](#). The total estimated amount of data is around 7GB, with datasets ranging from 10MB to 1000MB.

## INSANE - Data Management Plan

| <b>Work Package 1. Training and Management</b>  |        |        |          |
|---|--------|--------|----------|
| Type of dataset   | Origin | Format | Quantity |
| Research training on R (statistical computing and graphics)                                     | Report | PDF    | < 10 MB  |
| Initial teaching training course  | Report | PDF    | < 10 MB  |
| Research training on GIS (geographic information systems)                                       | Report | PDF    | < 10 MB  |
| Training on teaching teams (motivation, class management and innovation project preparation)    | Report | PDF    | < 10 MB  |
| Research training on LimeSurvey (survey development and data analysis)                          | Report | PDF    | < 10 MB  |
| Career Development Plan (with a description of the hands-on-training and management activities) | Report | PDF    | < 10 MB  |

| <b>Work Package 2. Insect Diversity (Research Objectives O1 and O2)</b>                    |                             |                               |          |
|--|-----------------------------|-------------------------------|----------|
| Type of dataset  | Origin                      | Format                        | Quantity |
| Collected insects lists  | Experimental                | dat/csv                       | < 100 MB |
| Visual censuses lists  | Experimental                | dat/csv                       | < 100 MB |
| Abiotic measurements   | Experimental                | dat/csv                       | < 100 MB |
| Statistical analyses   | Computational               | dat/csv                       | < 300 MB |
| DNA Barcoding  | Computational               | TAR.GZ (dat)                  | < 500 MB |
| Files containing graphical representations coming from experimental and computational data | Experimental, Computational | TAR.GZ (JPEG, PNG, TIFF, PDF) | < 500 MB |

## INSANE - Data Management Plan

| <b>Work Package 3. Human-nature interactions (Research Objective O3)</b>                   |                             |                               |          |
|--|-----------------------------|-------------------------------|----------|
| Type of dataset  | Origin                      | Format                        | Quantity |
| Questionnaires   | Experimental                | dat/csv                       | < 100 MB |
| Statistical analyses   | Computational               | dat/csv                       | < 100 MB |
| Files containing graphical representations coming from experimental and computational data | Experimental, Computational | TAR.GZ (JPEG, PNG, TIFF, PDF) | < 500 MB |

| <b>Work Package 4. Predictive tool (Research Objective O4)</b>                       |               |                        |          |
|--|---------------|------------------------|----------|
| Type of dataset  | Origin        | Format                 | Quantity |
| Diversity index (WP 2)   | Experimental  | TAR.GZ (dat, csv)      | < 200 MB |
| Wellbeing index (WP 3)   | Experimental  | TAR.GZ (dat, csv)      | < 200 MB |
| RGB Satellite images obtained from Google Maps                                       | Computational | TAR.GZ (JPEG, GEOTIFF) | < 200 MB |
| Diversity and wellbeing map  | Computational | TAR.GZ (JPEG, dat)     | < 500 MB |
| Deep Learning codes for creating diversity and wellbeing maps                        | Computational | TAR.GZ (Python, dat)   | < 500 MB |
| Codes for analyzing the correlation between diversity index and wellbeing index data | Computational | TAR.GZ (Python, dat)   | < 500 MB |

| <b>Work Package 5. Dissemination and Communication</b>  |                |                            |          |
|---|----------------|----------------------------|----------|
| Type of dataset   | Origin         | Format                     | Quantity |
| DMP (including updates)   | Report         | PDF                        | < 10 MB  |
| Final preprint versions of papers for expert audiences (Estimation: 5 preprints)                                    | Paper preprint | PDF                        | < 20 MB  |
| Presentations in international conferences for expert audiences (including talk presentations, videos, and posters) | Presentation   | TAR.GZ (PDF, Keynote, mp4) | < 500 MB |
| Outreach material (primary/secondary schools and the general public)  | Report         | TAR.GZ (PDF, Keynote, mp4) | < 500 MB |
| Presentations in internal seminars  | Presentation   | PDF, Keynote, mp4          | < 300 MB |

#### 1.4. Data value

INSANE aims to conceal urban development with the conservation of biodiversity and human wellbeing. Specifically, the project is focused on studying insect diversity and citizens' self-perception of happiness depending on the type of urbanization model where they live: land-sharing versus land-sparing. INSANE will also create a new predictive tool that will allow the estimation of these parameters in other cities just with images from Google Maps. The data supporting the project (scientific results and their dissemination) will reach a broad range of expert audiences: biologists, sociologists, urban planners, computer scientists, and professionals from urban development companies (*e.g.* urban designers, green areas managers).

INSANE will also generate data for pedagogical purposes that will reach different non-expert audiences such as primary and secondary students and teachers, undergraduate and graduate students (particularly in the broad fields of biology and sociology), and the general public.

INSANE project will be supported by the project data management plan. This includes the creation of technical reports describing the improvement of experimental protocols or statistical analyses (to obtain more accurate results) as well as a new data context management to maintain and extend the produced computational resources.

## 2. FAIR Data

### 2.1. Making data Findable

To ensure data visibility, the metadata system used for the description of the materials hosted in the UGR repository, DIGIBUG, is [Dublin Core Qualified](#). This is a metadata initiative adopted by the European repository [OpenAIRE](#). DIGIBUG assigns a unique identifier (handle) to each document and/or dataset, which allows the identification and citation of electronic documents.

Each dataset generated by INSANE will be recorded by a dataset identifier in the general file **INSANE\_DATASET\_LIST.pdf**, which will be hosted (and regularly updated) in [DIGIBUG](#). The information of each individual dataset will be included in a linked metadata file, which will be updated (if necessary) and will contain the information appearing in the table below.

| Contents of a generic Metadata file associated to a given Dataset |  |
|---|--|
| <b>Dataset Identifier</b>   | The ID will result from the naming convention provided in the next table   |
| <b>Title of the Dataset</b>                                       | The title of the dataset, which will be easily searchable and findable   |
| <b>Responsible Partner</b>  | Partner institution responsible for the creation of the dataset (always <a href="#">UGR</a> )  |
| <b>Work Package</b>   | Project work package ( <i>e.g.</i> WP2, see work package table in <a href="#">Annex A</a> )  |
| <b>Dataset Description</b>  | A brief description of the dataset ( <i>e.g.</i> <i>collected insect list</i> , specifying the location and date)                            |
| <b>Dataset Benefit</b>  | What are the benefits of the dataset ( <i>e.g.</i> the data will serve to reproduce some particular results and/or perform further analysis) |
| <b>Dataset Dissemination</b>                                      | Where will the dataset be disseminated ( <i>e.g.</i> peer reviewed journal)  |
| <b>Type Format</b>  | See table with formats in <a href="#">section 1.3</a> ( <i>e.g.</i> <b>dat</b> , <b>PDF</b> , <b>TIFF</b> , ...)                             |
| <b>Expected Size</b>  | Dataset size (see size in the work package tables in <a href="#">section 1.3</a> )   |
| <b>Source</b>   | How was the dataset generated ( <i>e.g.</i> experimental data)   |
| <b>Repository</b>   | <a href="#">DIGIBUG</a> (for final preprints we will also have an <a href="#">arXiv</a> version, <a href="#">section 3</a> )                 |
| <b>DOI (if known)</b>   | The DOI will be entered once the dataset has been deposited  |
| <b>Date of Submission</b>   | The date of submission will be added once the dataset has been uploaded on the repository  |
| <b>Keywords</b>   | Keywords associated with the dataset ( <i>e.g.</i> sociological surveys, DNA barcoding)  |
| <b>Version Number</b>   | Version number to keep track of changes ( <i>e.g.</i> V1.0)  |



## INSANE - Data Management Plan

The dataset identifier (first field in the previous table) will be created according to the convention presented in the table below. The corresponding metadata file will be named by adding **META** at the end of the identifier of its linked dataset file. All the metadata files will have extension **.dat**.

| Convention for creating the Dataset Identifier   |                         |
|--|-------------------------|
| Components   | Example                 |
| Project name   | INSANE (always)         |
| Two digits chronological number<br>(corresponding to the order of appearance in<br>INSANE_DATASET_LIST.pdf)  | 6                       |
| Title of the dataset   | Survey-on-urban-insects |
| Version of the dataset<br>( <b>DIGIBUG</b> will allow to keep several versions)  | V1.0                    |
| Work Package associated to the dataset   | WP3                     |
| Format of the dataset  | PDF                     |
| Example dataset file identifier: <b>INSANE_6_Survey-on-urban-insects_V1.0_WP3.PDF</b><br>Example metadata file identifier: <b>INSANE_6_Survey-on-urban-insects_V1.0_WP3_META.dat</b> |                         |

## 2.2. Making data openly accessible

The data hosted in the institutional repository **DIGIBUG** will be accessible to the complete research community. In this respect, the data generated by INSANE does not entail Intellectual Property (IP) rights such as patents, trademarks, or copyrights. The agreement, supervised by **UGR's Research Results Transfer Office**, between the beneficiary institution (**UGR**) and the DMP coordinator complies with the **IP-MSCA rules for access rights** (royalty-free basis) and results ownership. The data generated will be accessible through the **DIGIBUG** website and will be open to any user without restrictions.

In the table below we provide examples of software packages to open and/or use the data generated by INSANE for different operating systems (Ubuntu, MacOS, and Windows) and for all the formats. These examples are not unique and the potential user might use other standard packages and/or platforms. At least one of the options for each format is free software based (Ubuntu).

## INSANE - Data Management Plan

| <b>Format</b>          | <b>Examples of software (operating system) to open/use them</b>   |
|------------------------|---|
| <b>TAR.GZ</b>          | <b>TAR.GZ</b> package (Ubuntu), <b>tar</b> utility (MacOS), <b>7-Zip</b> (Windows)                            |
| <b>dat</b>             | <b>GNU emacs</b> (Ubuntu, MacOS), <b>Notepad</b> (Windows)  |
| <b>PDF</b>             | <b>Adobe Reader</b> (Ubuntu, Windows), <b>Preview</b> (MacOS)   |
| <b>JPEG, PNG, TIFF</b> | <b>gThumb</b> (Ubuntu), <b>Preview</b> (MacOS), <b>Adobe Photoshop</b> (Windows)                              |
| <b>Python</b>          | <b>Python sources releases</b> (Ubuntu, MacOS, Windows) in <a href="http://www.python.org">www.python.org</a> |
| <b>Keynote</b>         | <b>Keynote</b> (MacOS)<br>(another pdf version will be uploaded for each Keynote presentation)                |
| <b>mp4</b>             | <b>VLC</b> (Ubuntu, Windows), <b>MPlayerX</b> (MacOS)   |

### 2.3. Making data interoperable

INSANE aims to collect and document all the data generated in a standardized way ([Dublin Core Qualified](#)) to ensure that all datasets, which will be accompanied by the corresponding metadata file, can be interpreted and shared.

In this respect, a metadata file will be created and linked to each dataset. These metadata files will include all the information detailed in the first table of [section 2.1](#) (Contents of a generic Metadata file associated to a given Dataset).

## 2.4. Increase data re-use

The datasets will be made available for their re-use and be stored in [DIGIBUG](#) without any cost. If datasets are updated, the coordinator of this DMP will be responsible for managing the different data versions, making sure that the latest version is available. The policies adopted by [DIGIBUG](#) concerning licenses, availability periods, and quality are:

### 1. License.

All materials published in the [DIGIBUG](#) repository incorporate different licenses of the nonprofit organization [Creative Commons](#), in particular, the [Creative Commons 4.0 version of the Non-Commercial-ShareAlike CC BY-NC-SA license](#), which is the one recommended by [OpenAIRE](#). This license allows the reuse of data at the end of the project and the use by third parties.

### 2. Availability.

The data will remain reusable after the end of the project with no time limitations nor access restrictions, unless embargo or access restrictions are eventually indicated.

### 3. Quality.

The quality of the datasets is guaranteed by the [DIGIBUG](#) operating software, which performs routine backups and checking of the material hosted.

## 3. Allocation of Resources

All the datasets generated by INSANE will be allocated in [DIGIBUG](#) without costs (neither in the short nor in the long term), time limitation, or access restrictions. With this we will ensure and enhance [green open access according to the H2020 guidelines](#).

[Dr. Olivia Sanllorente](#), department of Zoology at the [UGR](#) (Spain), will be responsible for the data management within the INSANE project, in particular for the creation of the DMP and its subsequent updates, and for recording and updating the datasets generated by INSANE.

## 4. Data Security

The [Scientific Documentation Service of the UGR](#), located in the [Library of the Hospital Real](#) (Granada, Spain), coordinates the electronic management of the [DIGIBUG](#) repository. In particular, [DIGIBUG](#) incorporates a program for backup and preservation. In this respect, the INSANE datasets hosted in [DIGIBUG](#) will receive the same security treatment as the rest of the documents in this repository. All the responsibilities concerning data recovery and secure storage will go to the [Scientific Documentation Service of the UGR](#), which is in charge of the storing of the datasets hosted in [DIGIBUG](#).

## 5. Ethical Aspects

As stated in the Ethics Issues of the MSCA-IF-EF-CAR grant agreement (101033024 - 2020), the 4 requirements for ethical review of INSANE are already delivered (Requirements 1-4). However, this project does not involve the use of human cells or tissues, personal data collection and/or processing (surveys will be completely anonymous), animals susceptible of ethical aspects (insects do not need an ethical committee report), potential for misuse of research results, or elements that may cause harm to the environment, animals or plants.

## 6. References

- [1] [Guidelines on FAIR Data Management in Horizon 2020, version 3.0, 26 July 2016](#)
- [2] [Data Management Plan Template, UGR Library, version 1.0, 1 July 2017](#)
- [3] [J. Horst and C. Lynch, AMECRYS Data Management Plan, version 3.0, 29 March 2017](#)

## 7. List of Acronyms

| Acronym        | Meaning   |
|----------------|---|
| DMP            | Data Management Plan  |
| IP             | Intellectual Property   |
| MSCA-IF-EF-CAR | Marie Skłodowska-Curie Actions - Individual Fellowship - European Fellowship – Career Restart |
| UGR            | University of Granada   |

## 8. Annex A. Work Packages

The table below summarizes the main research objectives, tasks, training, management, and documents associated to each work package of the INSANE project (see tables in [section 1.3](#)).

|   |  |  |
|---|--|--|
| <b>Work Package 1.<br/>Management</b>   | <b><u>Title:</u> Training and Management</b>   | <b><u>Duration:</u> month 1 to 36</b>  |
| <p><b><u>Documents to be produced.</u></b><br/>Four Career Development Plan reports: one for the initial document and three updates. Three reports on protocols for field work (insect diversity: pit-fall, pan-traps, visual censuses). One report on protocols for sociological surveys. One report on lab protocols for genetic procedures. One report on statistical procedures. One report on protocols for satellite images selection. One report on the model for the supervised deep learning algorithm. One Final Report summarizing the results of the project.</p> |  |  |
| <p><b><u>Main Tasks.</u></b><br/>Hands-on training activities. Report writing. Elaboration of the sociological survey. Research training on statistical analyses. Research training on insect sampling and identification. Research training on sample preparation for genetic analyses.</p>  |  |  |
| <b>Work Package 2.<br/>Objectives O1 and O1</b>   | <b><u>Title:</u> Insect Diversity</b>          | <b><u>Duration:</u> month 5 to 24</b>  |
| <p><b><u>Research Objectives.</u></b></p> <ul style="list-style-type: none"> <li>• <b>Objective O1:</b> The effect of urbanization on insect diversity.</li> <li>• <b>Objective O2:</b> The effect of urban landscape organization on insect diversity.</li> </ul>  |  |  |
| <p><b><u>Main Tasks.</u></b><br/>Performing field work. Identification and preservation of samples. Elaboration of a reference collection of urban insects.</p>   |  |  |
| <b>Work Package 3.<br/>Objective O3</b>   | <b><u>Title:</u> Human-nature interactions</b> | <b><u>Duration:</u> month 5 to 24</b>  |
| <p><b><u>Research Objectives.</u></b></p> <ul style="list-style-type: none"> <li>• <b>Objective O3:</b> The effect of urban development (land-sharing or land-sparing) on human-nature interactions.</li> </ul>   |  |  |
| <p><b><u>Main Tasks.</u></b><br/>Performing citizens' surveys. Analyzing sociological data and integrating the biological data.</p>   |  |  |
| <b>Work Package 4.<br/>Objective O4</b>   | <b><u>Title:</u> Predictive tool</b>           | <b><u>Duration:</u> month 24 to 36</b> |
| <p><b><u>Research Objectives.</u></b><br/><b>Objective O4:</b> Development of predictive tool</p>   |  |  |
| <p><b><u>Main Tasks.</u></b><br/>Hands-on training activities. Report writing. Research training on machine learning. Research training on Geographic and Information Systems. Perform concept tests. Elaboration of codes for Deep Learning analyses.</p>  |  |  |

INSANE - Data Management Plan

|  |  |                                       |
|--|--|---------------------------------------|
|  |  |                                       |
| <b>Work Package 5.<br/>Dissemination</b>   | <b><u>Title:</u> Dissemination and Communication</b> | <b><u>Duration:</u> month 1 to 36</b> |
| <p><b><u>Documents to be produced.</u></b><br/>                 One DMP (with 2 repository updates). Five final preprint papers. Presentations for: three international conferences, three internal seminars, seven events for non-experts (outreach activities), material for research (reference entomology collection), educational sheets for primary/secondary students, good practice guide.</p> |  |                                       |
| <p><b><u>Main Tasks.</u></b><br/>                 Report writing. Repository updates. Paper writing. Preparation of material for conferences, field work, seminars, and events for non-experts.</p>  |  |                                       |
|  |  |                                       |