### **Proyecto**

Expansión del “Cortaderia selloan” en Cantabria

### **Participantes**

### Estela

### Diego

### Iker

### Gerardo

### Manuel

### **1. DATA SUMMARY**

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

*The purpose of data collection/generation is to build a geospatial map of the extension of the plant within the region of Cantabria.*

*This data collection/generation is related to the main objective of the project which consists in facilitating the decision making of the Cantabrian Government for the management and mitigation of the plant.*

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

*Types: Geospatial data (raster, coordinates), non-geospatial data (environmental, climate, plant genomic)*

*Format: OGC standards (GeoTIFF, NITF, JPEG, PNG, JPEG2000 and HDF), svg, tiff, float64, int64*

Will you re-use any existing data and how?

*We will use existing geospatial data from Cantabria. The first layer of geospatial data will be collected from the Cantabrian Government Web Coverage Service (WCS).*

What is the origin of the data?

* [Cartografía y SIG](http://www.territoriodecantabria.es/cartografia-sig)/
* [Servicios de Interoperabilidad OGC](http://www.territoriodecantabria.es/cartografia-sig/servicios-de-interoperabilidad-ogc)/
* [Servicios WCS](http://www.territoriodecantabria.es/cartografia-sig/servicios-wcs-iig)/
* <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=58286>

What is the expected size of the data?

*Geospatial data: Color, resolution 50 cm, 4 bands. JPEG 2000 Compression regular (80 Gb), JPEG 2000 Compression Optimized C (35 Gb), Geotiff 16 bits (160 Gb), Geotiff 8 bits (80 Gb). Total estimated: 355 Gb.*

*Non-geospatial data: plant genome (> 1 Gb).*

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

*This may be useful for policymakers, researchers, private companies specialized in mitigation of this type of plants.*

### **2. FAIR DATA**

#### **Making data findable, including provisions for metadata**

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

*Each dataset will be placed in a open datahub, using DOI as means of identification.*

What naming conventions do you follow?

*Snake case naming convention will be used throughout the project.*

Will search keywords be provided that optimize possibilities for re-use?

*Search keywords be provided that optimize possibilities for access and re-use.*

Do you provide clear version numbers?

*In terms of versioning, we will base on* [*Python's PEP 404*](https://www.python.org/dev/peps/pep-0440/)*.*

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

*Possible Metadata that will be created will come from many sites containing scientific information about the plant.*

*Our metadata will apply the good practices for Invasive Alien Species*

[*https://www.frontiersin.org/articles/10.3389/fams.2017.00013/full*](https://www.frontiersin.org/articles/10.3389/fams.2017.00013/full)

*Other common scientific names for Cortaderia Selloana:*

* *Arundo selloana Schult.*
* *Cortadeira argentea.*
* *Cortaderia dioica.*
* *Gynerium argenteum.*

*International Common Names:*

* *English: silver pampas grass; Uruguayan pampas grass.*
* *Spanish: cortadera; ginerio plumacho; hierba da las pampas; penacho.*
* *French: gynerion argente; herbe des pampas; roseau a plumes.*

*Local Common Names*

* *Brazil: bardeira; capim-do-pampas; paina; pluma.*
* *Germany: Silberweisses Pampasgras.*
* *Portugal: paina; plumas capim-das-pampas; ponacho-blanco erva-das-pampas.*
* *Spain: carrizo de la pampa.*

*Taxonomic Hierarchy:*

* *Domain: Eukaryota*
* *Kingdom: Plantae.*
* *Phylum: Spermatophyta.*
* *Subphylum: Angiospermae.*
* *Class. Monocotyledonae.*

#### **Making data openly accessible**

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

How will the identity of the person accessing the data be ascertained?

#### **Making data interoperable**

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

#### **Increase data re-use (through clarifying licences)**

How will the data be licensed to permit the widest re-use possible?

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

How long is it intended that the data remains re-usable?

Are data quality assurance processes described?

Further to the FAIR principles, DMPs should also address:

### **3. ALLOCATION OF RESOURCES**

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Who will be responsible for data management in your project?

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

### **4. DATA SECURITY**

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

*Nature has proposed a number of safe repositories. The following ones are recommended: Integrated Taxonomic Information System (ITIS) / KNB: The Knowledge Network for Biocomplexity / NCBI Taxonomy / Global Biodiversity Information Facility (GBIF) / Morphobank.org / Movebank Data Repository / Image Data Resource*

### **5. ETHICAL ASPECTS**

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

*No information coming from human participants will be collected. The genomic information about the “Cortaderia selloan” is public and can be accessed freely.*

*The cartographic pictures obtained by means of drones do not affect any sensible facility (such as militar or similar). In spite of this, these picture may include people that, based on the protection law of personal data, should not be identifiable. For this reason, persons will be pixeled in the pictures.*

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

*No personal data will be collected provided the above mentioned measures are applied.*

### **6. OTHER ISSUES**

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?