# RESUMEN Y TIPS EXTRAIDOS DE LOS EJEMPLOS PARA LA REALIZACIÓN DEL PLAN DE GESTIÓN DE DATOS

Contenido

[BLOQUE 1: DATA COLLECTION AND TIPO DE DATOS 2](#_Toc132729531)

[BLOQUE 2: METADATOS 8](#_Toc132729532)

[BLOQUE 3: ASPECTOS ÉTICOS, ALMACENAMIENTO Y SEGURIDAD 10](#_Toc132729533)

[BLOQUE 4: LONG TERM Y ACCESIBILIDAD 13](#_Toc132729534)

[BLOQUE 5: RESPONSABILIDADES 14](#_Toc132729535)

Este documento puede ser usado para elegir frases, conceptos y tips en relación con cada uno de los párrafos a incluir en el plan de gestión de datos. Elaborado por la unidad de metodología en el marco del taller de metodología a partir de los ejemplos ubicados en el directorio documentos de utilidad/Data\_Management\_Plans y que a su vez han sido creados utilizando las herramientas de creación de Planes de Gestión de Datos.

DMPTool <https://dmptool.org/users/sign_in>

DMPonline: <https://dmponline.dcc.ac.uk/public_plans>

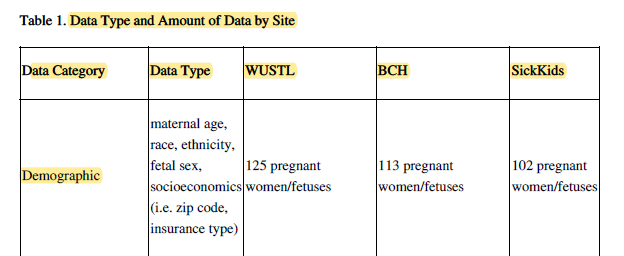
# BLOQUE 1: DATA COLLECTION AND TIPO DE DATOS

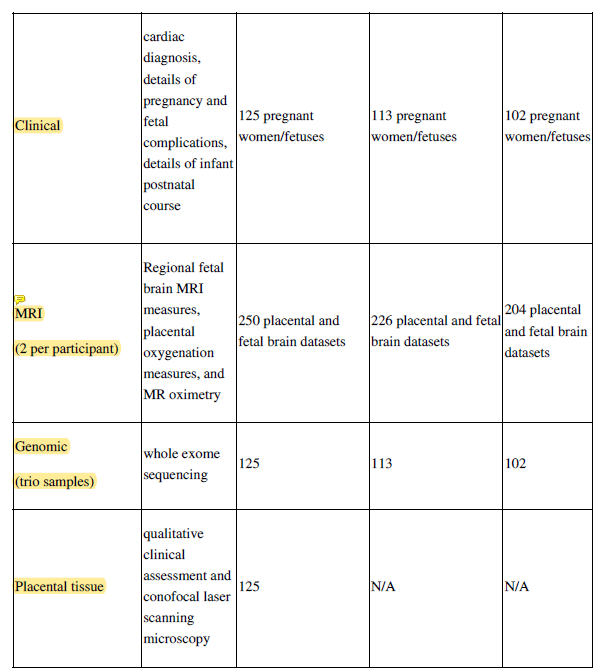
What data will you collect or create? How will be collected or created?

Categoría de datos por cada uno de los ficheros que se espera tener.

Tipo de datos (ejemplos de variables). Tamaño (nº especímenes, nº pac). Tipo de fichero (xls, csv, sav, image file tiff, jpg, etc): Tamaño esperado (1GB). Soporte (ej: Case Report Forms que luego se digitalizan).

**EJEMPLO 5 OBSERVACIONAL (placenta)**: The proposed project will involve human participants and will acquire demographic, clinical, fetal/placental magnetic resonance imaging (MRI), genomic, and placental tissue data from 225 pregnant women with a fetal diagnosis of CHD and 115 pregnant women with a healthy fetus.





**EJEMPLO 4- ECAs**

**How will data be collected, created or reused?**

Data regarding participants/families will be collected by trained professional healthcare staff in the participating clinics according to the study protocol instructions, utilizing pre-defined Case Report Forms and pre-defined variables.

Recruitment, eligibility assessments, inclusion and randomization of research participants at pre-defined time points is integrated in the clinical care at the cancer genetic clinics involved in the study (4 sites in Sweden).

Most documentation will be done by the assigned research nurse/collaborating staff at each unit who is conducting the genetic counselling with the participating patients, and research data is handled with the same security level as any other patient data on the clinic before being handed over to the national secretariat.

**What types of data will be created and/or collected, in terms of data format and amount/volume of data?**

The four cancer genetic clinics involved in the study will collect and document necessary data points related to the study outcomes. Our research study will collect:

1. **Personal patient data** including name, birthdate, local clinical identifier, address, preference on communication channel and compliance with each of our inclusion criteria. During the recruitment medical data is added once available – i.e results from

genetic diagnosis.

2. **Outcome data** in the form of family data and genetic investigation results as well as information about number of at risk relatives for each study participant, and whether these have made contact with a cancer genetic clinic at the time of outcome

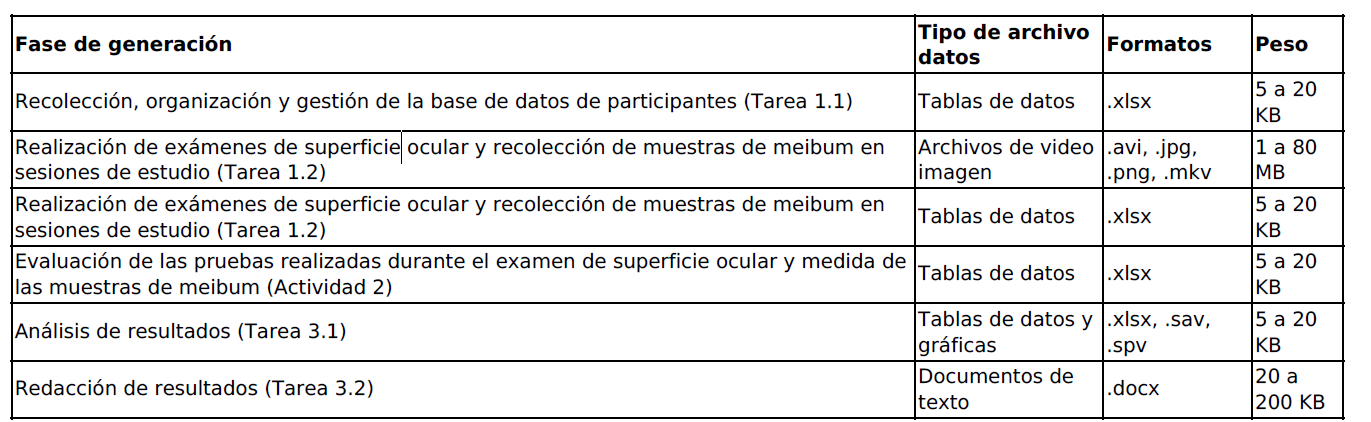
summary (T=12).

3. **Patient Reported Outcome Measures** regarding wellbeing, stress, anxiety etc. (through validated instruments at time points T=0 and T=6 months following initial risk disclosure to the participant.

4. **Screening log and inclusion log** will be kept locally as temporary documents to enable invitation, recruitment and inclusion.

The research database designed in the Access Software environment will contain transferred variables from CRF1 and CRF3 as well as questionnaire data to enable analysis, sub-group analysis and comparison of study groups in regard of the pre-defined outcomemeasures. The Research Database will only hold variables in relation to the specific individual study participant number (for personal identification one would need the data key kept locally at the study clinics or in the raw data CRF1). These data keys will be destroyed as soon as the analysis in the project is complete, checked and verified (current estimated storage 15 years).

**Ejemplo 6 OJO seco correlación**



**Ejemplo 9. Cuestiones a tener en cuenta (RECOGNITION) OMICs**

**1.3 In collecting data for my project, I will do the following:**

Add new data to an existing data set (please specify)

Generate new data

Use existing data (please specify)

**1.4 In my research, I will use:**

A combination of quantitative and qualitative data

Quantitative: RNA-seq data, proteomics data, most clinical outcome measures.

Qualitative: gene / transcript / protein annotations.

**1.8 I can give an estimate of the size of the data collection; specifically, the number of participants or subjects (“n=”)**

**in the collection and its size in GB/TB**

Yes (please specify)

RNA-seq:

N=30 participants x 2 time points (baseline and after 10m of intervention)

Average size of raw RNA-seq data per sample: 10 GB

Average size of analysed data per sample: 20 GB

Total: 30x60 GB = 1.8 TB

N=256 participants x 2 time points (baseline and after 10m of intervention)

Average size of raw proteomics data per sample: 10 GB

Average size of analysed data per sample: negligible

Totaal: 512 x 10 GB = 5.2 TB.

**1.9 The following end products I will make available for further research and verification (please elaborate briefly)**

Syntaxes

Documentation of the research process, including documentation of all participants

Data documentation

Biobank

(Several versions of) processed data

Raw data

Raw and processed RNA-seq data and metadata in European Genome Archive (EGA) (data access committee in place and data

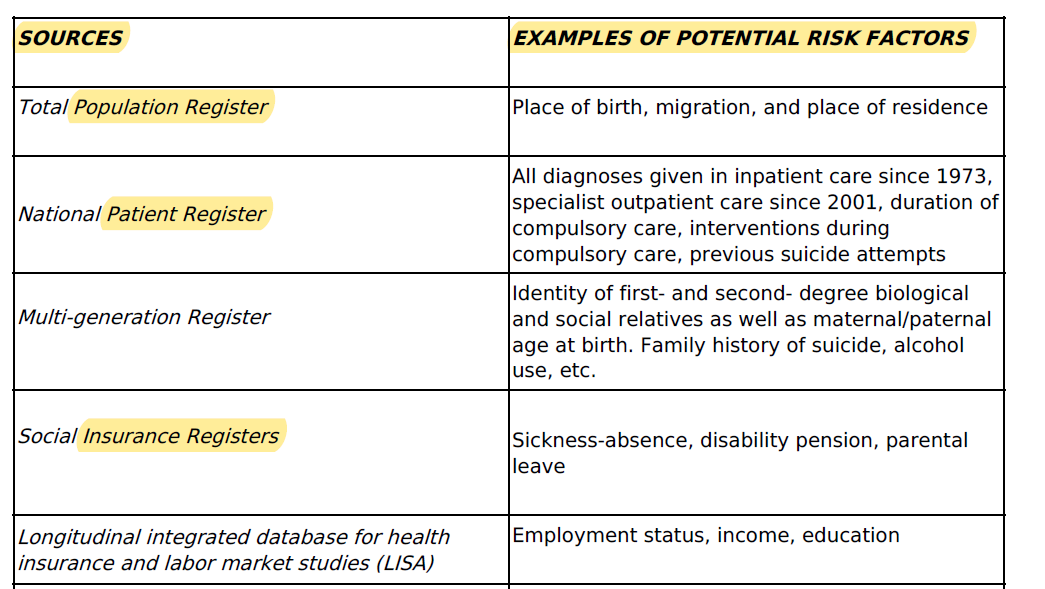
transfer agreement in place).

Proteomics data and metadata in PRIDE repository at the European Bioinformatics Institute (to be done)

Biobank samples: via Eurobiobank at two locations (Newcastle and Munich).

Documentation, syntaxes in github.

**Ejemplo 1. RWD suicidio**



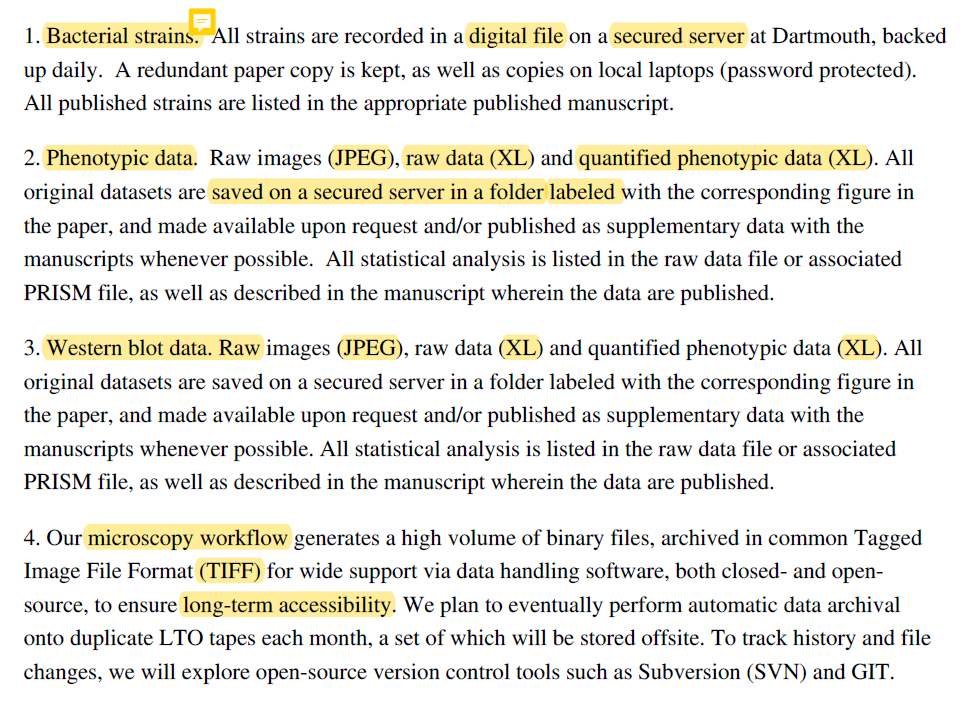
7 BASICO Alzheimer

Data generated via the following methods: flow cytometry, confocal microscopy, fluorescence microscopy, ELISA, and absorbance measurements.

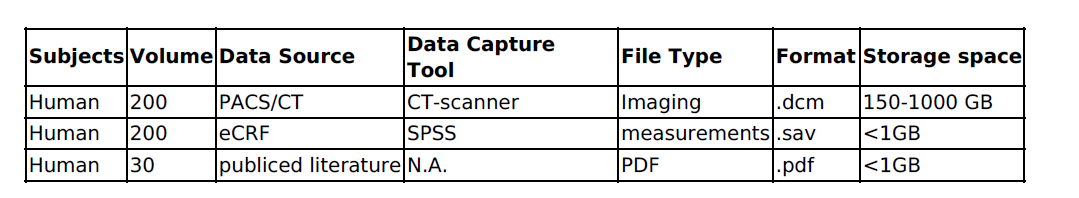
This data will be collected from a minimum of 3 independent in vitro experiments, with each independent experiment consisting of mutiiple group comparing different SHIPi approaches The total size of the data collected is projected to be 50 GB.

We expect to generate the following data file types and formats during this project: Carl Zeiss microscopic image file (.CZI), images (.TIFF), tabular (.CSV), flow cytometry list mode data. Raw data files will be analyzed to generate Prizm or Excel files containing replicate measurments in the above to enable statistical analysis.

**BIOFILMS**



OTRO (NO SÈ CUÁL):



**10 Cualitativo profilaxis HIV Delphy**

This project will produce qualitative and quantitative generated from formative semi-structured interviews, surveys, and electronic health records. Qualitative data will be collected from 75 participants in the formative phase and 20 participants during exit interviews, generating two datasets. Quantitative survey data will be collected from 38 participants in the modified Delphi consensus building and 20 participants in the longitudinal survey of clinic providers and staff, generating two datasets. Electronic health record data will be collected from clinic patients to

generate a single dataset.

Raw data will be de-identified and transferred to MAXQDA, SPSS, STATA, and/or Mplus for analysis. To protect research participant identities, de-identified individual qualitative and quantitative data will be made available for sharing. Summarized electronic health record data will only be made available for sharing. In circumstances where small sample sizes may easily allow reidentification

of research participants, only summarized data will be made available for sharing.

We expect to generate the following data file types and formats during this project: text (.docx, .txt), and quantitative data (.dat, .sav, .csv). The total size of the data collected is projected not to exceed 1 GB.

**11 Parkinson disease**

What data will you collect or create?

Socio-demographic characteristics of the patients: age, gender, academic level, socio-economic level, as well as a clinical anamnesis of the same (pathologies, symptoms and associated pharmacology).

Aspects intrinsic to the immersive Virtual Reality program:

Security: It will be tested using the Simulator Sickness Questionnaire. (Kennedy, 1993)

Usability: It will be tested using the system's usability scale. (Brooke, 1996)

Personal Experience: It will be tested through Game Experience Questionnaire-post-game.(Ijsselsteijn, 2013)

Adherence to the program: An ad hoc record sheet will be executed.

Functional and quality of life aspects:

Balance, gait and risk of falling: It will be evaluated using the Tinetti Test. (Kegelmeyer, 2007)

Functional balance: It will be evaluated through the “Five Times Sit to Stand test-FTSTS. (Duncan, 2011)

Functional autonomy: it will be sent through the Timed Up and Go test with the WIVA application. (Mollinedo et al., 2018)

Symptoms and Follow-up in the progression of Parkinson's disease: It will be presented through the (MDS-UPDRS) (Rodríguez-Violante, 2012)

Quality of Life through the PDQ-8 questionnaire (Jenkinson et al. 2007)

Monitoring and control of sessions. Due to the lack of previous experiences in the application of this type of immersive virtual therapies in the population diagnosed with Parkinson's, it is necessary to carry out a control and monitoring during each of the sessions of the InViPark program in order to identify possible contraindications. For them we will apply the following

tests and trials:

Safety: It will be evaluated through the Simulator Sickness Questionnaire. (Kennedy,1993)

Usability: It will be evaluated using the System Usability Scale. (Brooke, 1996)

Personal Experience: It will be evaluated through Game Experience Questionnaire-postgame.(Ijsselsteijn, 2013)

Adherence to the program: An ad hoc registration sheet will be made o Effort made: Youwill be monitored through a wrist-based heart rate monitor.

Perception of Effort: It will be evaluated through the Modified Borg Scale. (Sousa, 2016)

Performance: Score given by the system in the activity carried out.

How will the data be collected or created?

The data will be collected through a computer application on (limesurvey) and will be stored in the server that the work team has at the Faculty of Education and Sports Sciences. Two folders will be created in which two files will be saved at the beginning, one with the personal and identifying data of each patient (InViPark\_Iv1) and another with the data to be evaluated that will always be encoded

(InViPark\_Dv1). Each time any of the files is updated, the version will be changed. Each folder will have two odd / even subfolders in which the file will be saved depending on the odd or odd version. The database created by limesurvey will be imported into STATA or SPSS, for data analysis. Each

time the database is updated, the PI of the project, the person in charge of the database, will perform a frequency analysis to identify possible errors in the introduction of the same.

# BLOQUE 2: METADATOS

**(Y en algunos formatos DATA QUALITY)**

**What documentation and metadata will accompany the data?**

**(otros similares). How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, etc.?**

Se creará un diccionario de datos (*o, si procede, un modelo de datos, a valorar*) (o un README.txt) en el que se recoge el listado de variables, la naturaleza de cada variable, el estándar utilizado en su caso y las versiones del mismo. Si procede: Dicho modelo de datos se publicará en ZENODO.

Ejemplo modelo de datos en zenodo:

<https://doi.org/10.5281/zenodo.7778291>

EJ1

Documentation of the material follows the approved guidelines of the Department of Medical Epidemiology and Biostatistics (MEB) at KI. These include a standardized folder structure for documentation comprising of Codebooks (metadata about data), Logbooks (metadata about data processing and cleaning), Analysis plans (including detailed descriptions of the data retrieval and research studies), Manuscripts, Syntax scripts and output files from database systems and statistical software (for data management and analysis), Program folders, Data folders and Communications with data providers. The Department also has a standard for variable naming and coding for primary data collections. Every authority has their own detailed description of the specific register data. A general description of all gathered registers can be found on the research group’s server. The usage of the data will be documented with the Karolinska Institutet’s (KI) Electronic Lab Notebook (ELN). Every researcher is required to give detailed documentation of how and what the data is used for, additionally also write a detailed description plan before the start of every research question within the project.

EJ 2: standards

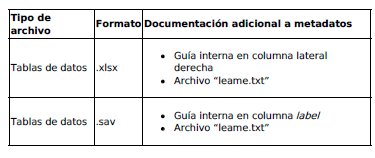
Data will be stored in common and open formats, such as csv for our in vivo study data. Information needed to make use of this data[e.g. the meaning of variable names used as column headers, information about missing data if any along with references to the sourcesof those standardized names and metadata items will be included wherever applicable.

EJ5

To facilitate the interpretation and reuse of the data, a data dictionary will be generated and deposited into a repository along with all shared datasets. The data dictionary will define and describe all variables in the dataset. A README file will also be generated and deposited into the repository and will include instructions for data analysis.

EJ 6

El presente proyecto no contempla la generación de archivos de metadatos durante su realización. En aras de interpretación de hojas de datos para el análisis estadístico tipo “.xlsx” o “.sav”, estas van acompañados de un código para la interpretación de acrónimos dentro de una columna del propio documento (documento tipo “.xlsx”) o en las columnas de label (documento tipo “.sav”); en los casos de alto manejo de acrónimos o códigos en listas de datos que requieran de interpretación, el documento de datos va acompañado de un documento tipo “leame.txt” con la clave de interpretación necesaria.



Ej 8

This document describes the initial Data Management Plan (DMP) for the RESETageing project, funded by the EU’s Horizon 2020 Programme under Grant Agreement number 952266. This document defines the general policy to data management in RESETageing project. This DMP allows the alignment of RESETageing consortium regarding the data management generated during the project. Isintended to be a guide on how partners should generate, acquire, handle, share and curate their research data within RESETageingproject.

The DMP is a living document and will be updated throughout the project whenever significant changes arise.The project is in its initial phase and therefore at this stage it is not yet possible to anticipate all the types of data that will begenerated during RESETageing project development.The template for the description of the datasets is available to each partner of RESETageing consortium in order to elaborate the detailed description for each dataset generated within RESETageing project. As the datasets are identified they will be described and added in the future versions of this living document. The elaboration of RESETageing general data management policy has been developed in accordance to the open data requirements of Horizon 2020 FAIR principles and it follows the template for Horizon 2020 FAIR Data Management Plan (version 3.0). It applies mainly to new results that are produced in RESETageing and that are to be made available by the project consortium as open source, open science and open data.

Ej 10

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To facilitate interpretation of the data, interview guides, survey measures, and questionnaire files will be created, shared, and associated with the relevant datasets. Documentation and support materials will be compatible with ClinicalTrials.gov Protocol Registration Data Elements.

# BLOQUE 3: ASPECTOS ÉTICOS, ALMACENAMIENTO Y SEGURIDAD

**How will you manage any ethical issues? How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?**

**How will the data be stored and backed up during the research?**

**How will you manage access and security?**

Describir que todos los datos serán pseudonimizados. Describir cómo se va a garantizar dicha seguridad en la protección de la información (uso de contraseñas, servidores de seguridad, etc).

Aquí se podría incluir el párrafo de aspectos éticos.

Incluir los apartados del ejemplo IdiSNA

EJ 1:

Extraction of register data has been done by the relevant agency (e.g., Statistics Sweden extracts, merges, quality checks, and handles the Swedish Labor Market Data (LISA)), the Swedish data registration, storage and its validation of input and quality is

documented. At the merging and management of different register data in the research group, detailed forms are filled in by each researcher who then work closely with a statistician or data manager to get the required variables. Further management of data is documented by the relevant researcher in KI’s ELN.

**Legal and ethical aspects**

**How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data,confidentiality and intellectual property rights?**

KI as an organization complies with GDPR in both legal and ethical aspects. More information can be found at <https://medarbetare.ki.se/gdpr> All identifying information such as name and personal identity number have been removed before data sets are made available to personnel for analysis; we only work with pseudonymised data.

**How is correct data handling according to ethical aspects safeguarded?**

Ethical permit for the project has been granted by the Regional Ethical Committee in Stockholm (protocol number: 2013/862- 31/5;2020-06540). The present study will exclusively use already gathered data from nationwide registries which is extensively usedat KI. These data will only be handled in pseudonymised form and kept on the safe server at MEB at KI protected by doubleauthentication and other modern research standards to ensure participants safety and integrity. The key for breaking

pseudonymisation is not and will not be available to the researchers as it is kept by the appropriate government agency (Socialstyrelsen) which routinely performs this safekeeping for Swedish registry research.

EJ 4: We will handle sensitive personal data according to GDPR-regulations, on the basis of performing a state delegated mission AND based on informed consent by participants in the study. All access to medical records is done within the clinical setting, by the healthcare staff involved in the care of the patient, and only data in pre-defined structured case report forms (CRF’s) are relayed to the research team.

All eligible patients are invited to the study with written and spoken information and consent forms are collected for each participant. The information and consent form (forskningspersoninformation, FPI) is approved by the National Ethical Review Authority (Etikprövningsmyndigheten) and includes required items about patients’ rights to access collected data, withdraw from the study at any time, possible hazards, future dissemination or research results etc.

EJ 6\_

**How will you manage any ethical issues?**

Todos los integrantes del proyecto de investigación son titulados universitarios según los requisitos establecidos en la Orden Ministerial CIN/727/2009 del 18/03/2009 que habilitan para el ejercicio de Óptico-Optometrista (BOE nº 73 de 26/03/2009), por lo que todas las pruebas clínicas realizadas en el presente proyecto están avaladas y son ejecutadas por profesionales colegiados cuya formación les permite ejercer bajo las competencias establecidas en la Ley de ordenación de las profesiones sanitarias 44/2003 del 21 de noviembre, Artículo 2.2 epígrafe b del Título Preliminar y en el artículo 7.2 epígrafe e del Título II. Así mismo, los principios de los estudios y protocolos que conforman el presente proyecto están diseñados de acuerdo con los principios de la Declaración de Helsinki. Los procedimientos del proyecto están sustentados en la legislación vigente de la Unión Europea, del Estado Español y de la Comunidad Autónoma de Galicia, así como en las normas de “Buena Práctica Clínica” y Código Deontológico que desde el punto de vista legal, ético y metodológico amparan los derechos individuales de las personas que voluntariamente participan en la presente investigación clínica. Todo aquel participante en el presente proyecto de investigación debe firmar un Consentimiento Informado específico para el proyecto, en el que se les indicará de manera concreta y detallada el tipo de Intervención Investigadora, Procedimientos y Protocolo, especificando así mismo su duración. Cualquier participante es libre de cesar su participación si así lo desea sin necesidad de explicación. En lo referente a los requisitos éticos o legales del presente proyecto, todas las actividades realizadas bajo el marco del Servicio de Optometría de la Universidade de Santiago de Compostela se encuentran aprobadas por el Comité de Ética de la Investigación de

Galicia (Santiago-Lugo, código de registro 2013/360). Además, el Comité de Bioética de la Universidade de Santiago de Compostela ha aprobado el desarrollo los siguientes estudios de investigación enmarcados dentro del presente proyecto: “Avances en la Evaluación del Síndrome de Ojo Seco Evaporativo. Disfunción de las Glándulas de Meibomio” con el código de registro USC-40/2020 y “Evaluación de los cambios en los síntomas y signos de los pacientes con Ojo Seco tratados con diferentes lágrimas artificiales e

higiene palpebral” con el código de registro USC-08/2021.

COPYRIGHT

(en muchos casos no procederá o no se menciona)

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Durante el proceso de realización del proyecto los derechos de autor y los derechos de propiedad intelectual de los datos que recopilan o crean pertenecen a los investigadores del Grupo de Investigación en Optometría (GI-2092 en el catálogo de grupos de investigación de la USC) siendo su principal propietario la directoria del mismo grupo (la Dra. Eva Yebra-Pimentel Vilar. Así mismo, no se contempla la reutilización de datos previos o de estos mismos datos una vez finalizados los protocolos del proyecto y cumplidos los objetivos de difusión y publicación marcados durante el mismo. De la misma manera, estos datos no serán cedidos a terceros, ni se establecerá o propondrá la creación de licencias de uso de los datos o patentes generados de los hallazgos obtenidos. En caso de algún cambio generado por circunstancias no contempladas durante la planificación del proyecto, siguiendo lo establecido por la “Orden por la que se aprueba la convocatoria de tramitación anticipada para el año 2022 del procedimiento de concesión de ayudas a «Proyectos de Generación de Conocimiento» y a actuaciones para la formación de personal investigador predoctoral asociadas a dichos proyectos, en el marco del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021- 2023” en su artículo Articulo 23.3.a, los integrantes del grupo de investigación se asegurarán de tomar las medidas oportunas que permitan conservar los derechos de propiedad intelectual necesarios para dar cumplimiento a los requisitos de publicación en revistas de acceso abierto. Así mismo y ante circunstancias necesaria, se establece la posibilidad de tomar las medidas oportunas para proteger, con carácter previo a la publicación científica, los derechos sobre los resultados de la actividad de investigación, desarrollo e innovación, de acuerdo con las normativas nacionales y europeas en materia de propiedad intelectual e industrial, obtenciones vegetales o secreto empresarial; en caso de cambio en los citados términos, se actualizaría así mismo en presente PGD.

**STORAGE**

**Which data are of long-term value and should be retained, shared, and/or preserved?**

P1:

**How is storage and backup of data and metadata safeguarded during the research process?**

Access to storage of data is guarded strictly by IT-policy at the Department with different levels of authorization given to a user (researcher/non-researcher) on the PI’s approval. The Department’s research data and other storage are backed up every day with snapshots of different versions available to recall.

**How is data security and controlled access to data safeguarded, in relation to the handling of sensitive data and personal data, for example?**

The present study will exclusively use already gathered data from nationwide registries which is extensively used at MEB. These data will only be handled in pseudonymized form (individuals will not be identifiable) and kept on the safe server at MEB at KI protected by double authentication and other modern research standards to ensure participants safety and integrity. All archived data in the project will beheld strictly confidential with pseudonymized data, and stored in the security server at the department. Everyone working with the data is required to attend a course "Good Data Management Practice".

P3:

Repository where scientific data and metadata will be archived: Provide the name of the

repository(ies) where scientific data and metadata arising from the project will be archived;

see Selecting a Data Repository)

The UCLA facility uses a primary data server with redundant storage. Raw data is routinely backed

up to external hard drives (cold storage), with multiple data servers at different locations to mirror

the data. At Dartmouth, all original datasets are saved on a secured server (DartSF).

For UCLA and Dartmouth, all original datasets and the resulting figures/tables are placed in a folder

labeled with the corresponding figure in the paper (i.e., using the format recommended by the

journal eLife), and made available upon request and/or published as supplementary data with the

manuscripts. The secured server at Dartmouth is backed up daily, off site. Original datasets are also

kept on password-secured laptops and deposited, whenever possible on journal databases.

Bulk cyclic nucleotide measurements are performed at the Michigan State University Metabolomics

Core (Text provided by the MSU Core Facility). All instrument raw data is identified by date,

investigator, experiment info and sample number and backed up to a Sharepoint server at MSU.

Processed data/tabulated results are likewise identified and backed up to the MSU Sharepoint server.

Processed/tabulated data are shared with the investigator and raw data files are also shared with

investigators upon request (these data files are handled as described here). Raw data files are also

converted to open-source (vendor independent) formats and uploaded with corresponding

experimental metadata to public data repositories such as the Massive DB or Metabolomics

Workbench upon request.

Laboratory notebooks and PCs will always remain the property of the institutions, with copies given to graduating investigators to facilitate future interactions.

All data will be also deposited into Zenodo, using the account geiselbiofilm.

The code for processing and analyzing data are stored in repositories with Github.

# BLOQUE 4: LONG TERM Y ACCESIBILIDAD

**What is the long-term preservation plan for the dataset?**

**How will you share the data? Are any restrictions on data sharing required?**

**How, when and where will research data or information about data (metadata) be made accessible? Are there any conditions, embargoes and limitations on the access to and reuse of data to be considered?**

How will the use of unique and persistent identifiers, such as a Digital Object Identifier (DOI), be safeguarded? Karolinska Instiutet has a central database with DOIs of all the published articles, which is backed up regularly.

Explicar qué se va a hacer para que los datos sean findables, accesibles y entendibles. Decir durante cuántos años se espera garantizar acceso.

Explicar qué datos van a ser guardados ese tiempo (tanto soporte físico y cómo y donde como servidores seguros del material digitalizado)

P1:

Since data is subject to GDPR, only metadata from registers can be accessible by anyone, due to the open metadata information on each authority’s webpage. The register data that we have gathered can only be used by the research group at KI with the purpose that the group has been granted the ethical approval for.

**In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be**

**made?** When the project is completed, data archiving will be done according to KI’s rules and policies. The archiving guidelines include instructions for selection of files necessary to ensure reproducibility of published results, as well as safeguarding the use and readability of valuable data for future research.

P2:

There are no anticipated factors or limitations that will affect the access, distribution or reuse of the scientific data generated by the proposal. Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Controlled access will not be used. The data that is shared will be shared by unrestricted download.

P4:

**In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?**

All relevant research data generated in the DIRECT-project will be stored as required by law within the appropriate IT-environment recommended by Umeå University IT-services at the end of the project period. Analog paper data (such as CRF's and consent forms) will be stored in secure archiving facilities associated with The Department of Radiation Sciences, Umeå University.

**Will specific systems, software, source code or other types of services be necessary in order to understand, partake of or use/analyse data in the long term?**

Only standard easy-access file formats will be used for the digital long term storage of research data in order to facilitate access if needed in the future.

**How will the use of unique and persistent identifiers, such as a Digital Object Identifier (DOI), be safeguarded?**

If applicable, the repository publishing our metadata will automatically assign a Digital Object Identifier (DOI).

# BLOQUE 5: RESPONSABILIDADES

Who will be responsible for data management?

P4

Responsibility and resources

Who is responsible for data management and (possibly) supports the work with this while the research project is in progress? Who is responsible for data management, ongoing management and long-term storage after the research project has ended?Barbro Numan Hellqvist is in charge of the databases and overall data management. Dan Harnesk has the responsibility for overall information security at Umeå University.

What resources will you require to deliver your plan?

What resources (costs, labour input or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)? What resources will be needed to ensure that data fulfil the FAIR principles?

The project funds a part-time data manager statistician for the running of the project. Additional staff resources for operational management of the project will provide resources to fulfil the fair principles - in this case regarding metadata from our study.

Financing of necessary IT solutions for secure storage and archiving of databases and raw data materials are covered within thebudget of the project.