

# Neurorradiología e Inteligencia Artificial

#CSIC



*Special thanks to Lara Lloret,  
David Rodriguez, Fernando Aguilar,  
Alvaro López from IFCA*

## INVESTIGACION Y BIG DATA

**Prof. Jesús Marco de Lucas**  
*(jesus.marco [at] csic.es)*

*Vicepresidente de Investigación Científica y Técnica*  
**Consejo Superior de Investigaciones Científicas,  
CSIC, SPAIN**

*October 26th 2019 @Santander*

*Disclaimer: all statements presented are based in the personal  
experience as researcher and do not reflect any institutional position*



#CSIC

# Understanding the research ecosystem

*Who eats whom (what is the biomass? money? knowledge!)*

**SCIENCE IS IN THE BUSINESS CORE IN OUR CENTURY !**

*Science is no longer only the main interest of scientists...*

GOVERNMENT AND PUBLIC ADMINISTRATION

PRIVATE AND PUBLIC COMPANIES

PUBLIC RESEARCH FUNDING & PERFORMING ORGANIZATIONS

- research centers
- medical centers
- technology transfer offices
- students
- citizens

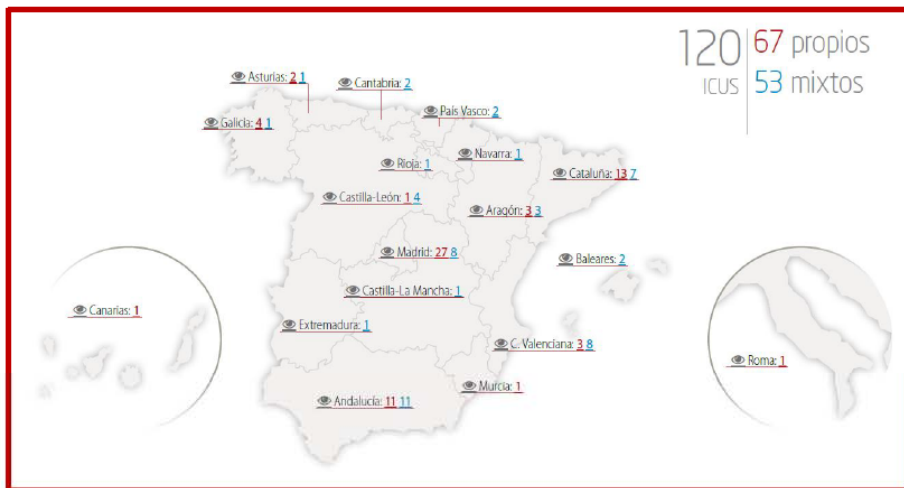
**SOCIETY**

*“Publishing journals as a **profit-maximizing business** is certainly as legitimate as it is for other **distributors of digital content based on intellectual property protections**. The research enterprise and its stakeholders are responsible for the future of scholarly communication”  
From “**Open Science by design**” NASEM*



# Consejo Superior de Investigaciones Científicas (CSIC)

#CSIC



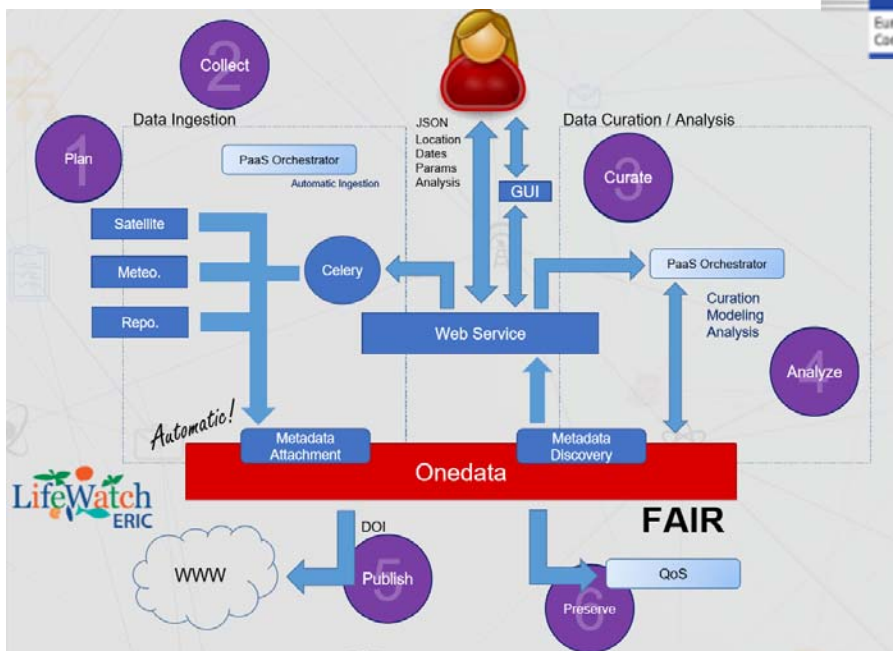
INDICADORES ANUALES: >13.000 artículos / >1.600 contratos / >125 patentes solicitadas

<b>10.642 Empleados</b>	<b>3.644 Investigadores</b>	<b>1.263 Investigadores en formación</b>	<b>4.472 Personal de apoyo a la investigación</b>	<b>1.263 Gestión</b>
5.220 mujeres	2.285 hombres	654 mujeres	2.445 mujeres	761 mujeres
5.422 hombres	1.360 mujeres	609 hombres	2.027 hombres	502 hombres

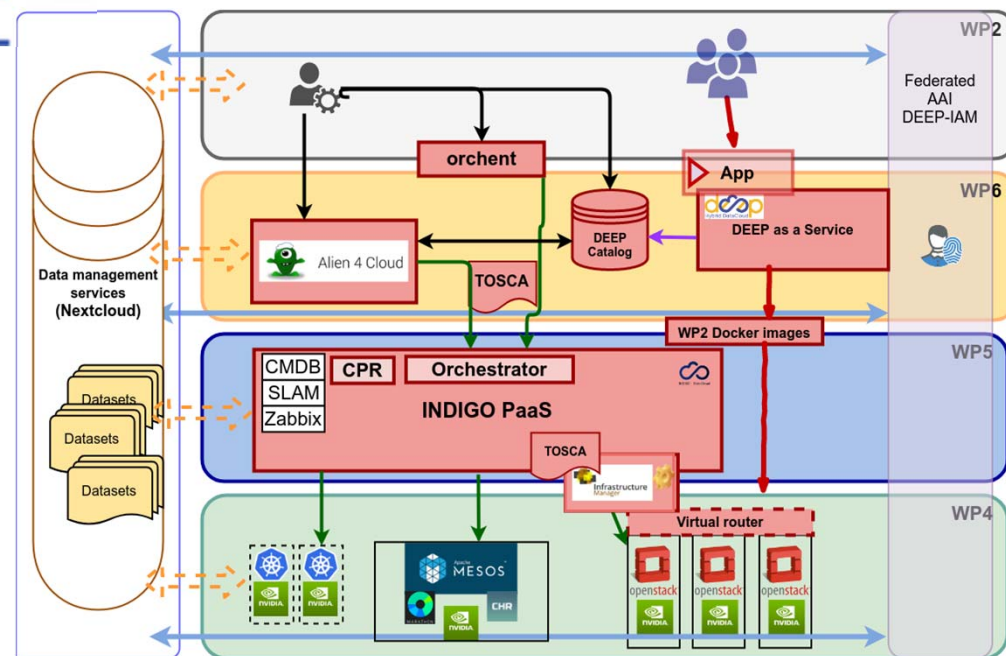


# Research Data Life Cycle and Cloud Platforms

#CSIC



F. Aguilar, XDC H2020 project



A. López, DEEP Hybrid DataCloud H2020 project

implemented following

**A set of Common Software Quality Assurance Baseline Criteria for Research Projects**

<https://github.com/indigo-dc/sqa-baseline>



## Examples: understanding data and AI for research

### Application to biology and medicine (examples)



- Successful projects developed on classification of plants and snails using a CNN.
- Some projects being started right now on CNN for medical imaging applications:
  - Mental diseases through nuclear magnetic resonances.
  - Fatty liver: Classification of liver biopsies
  - Brain hemorrhage through TAC images.



Exploiting same AI techniques:  
From plants to plankton, to conus, to HEP collisions, to brain...



Ignacio Heredia

**Arquitectura**  
ResNet50  
**Framework**  
Python con Lasagne/Theano  
**Training dataset**  
PlantNet (6K especies | 250K imágenes)

How to become the BEST expert in plant identification in your country in 10 minutes



**•Training dataset**  
Colección de imágenes de expertos (68 especies | 1.5K imágenes) que cubren tres regiones diferentes:

- Región Panámica
- Región de África del Sur
- Atlántico Occidental y Mediterráneo

**•Resultados**  
Los resultados usando sets de imágenes de Google son prometedores

<http://conus.deep.ifca.es/>

Do you think experts understand how they identify "by heart"?

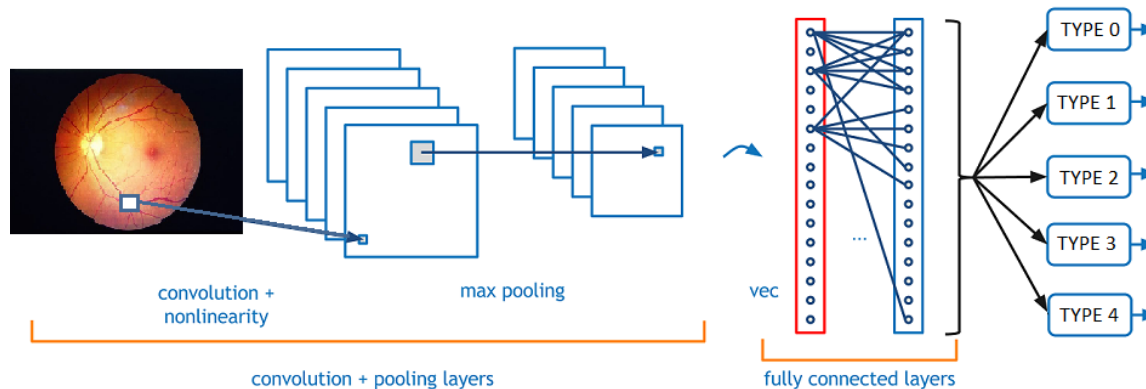
Lara Lloret



# Research in Medicine: Joining Data and AI

#CSIC

- **Retinopathy** is a fast-growing cause of blindness over 400 million people at risk from diabetic retinopathy alone : successfully treated if it is detected early.
- Colour fundus retinal photography is analysed in order to document the presence of disorders and monitor their change over time.
- Specialized medical experts interpret such images and are able to detect the presence and stage of retinal eye disease such as diabetic retinopathy.
- **Deep Learning approach:** automated classification based on color fundus retinal photography images



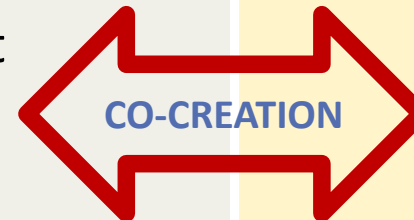
**Only an example** (but one of those few already accepted ) of **medical applications** where AI **should** help for good!

ECOSYSTEM:

- biomed researchers /
- doctors
- patients
- biomed companies
- authorities
- hospital owners
- assurance companies
- families

## A proposal: Joining Data and AI in a public safe environment

- **Donation of Medical/Personal Data**
  - Clinical History
  - Activity, Exercise, Food, Environment
- **Data Safe Haven (+ HPC framework)**
- **Legal Sandbox**
  - Joint (Public-Private) Research proposals
  - Open Research Software Tools (on IA)
  - Joint (Public-Private) Exploitation



ECOSYSTEM:

- biomed researchers /
  - doctors
  - patients
- biomed companies
  - authorities
  - hospital owners
- assurance companies
- families

Many, many technical advances

Ecosystem: **PRIVATE-PUBLIC COLLABORATION**  
**ON INTERDISCIPLINARY COMPLEX PROBLEM**

Key factor: **PRIVACY**

Status: **HOW TO PROGRESS?**

**EXPLOIT MEDICAL DATA ON**  
**“PUBLIC” HPC CLOUD?**  
**(IBM, Amazon, Google, Microsoft...)**

# What next from CSIC

- #CSIC**
- **Better understand each ECOSYSTEM** in Open Science, and our role
    - NOT ONLY MONEY BUT KNOWLEDGE (AND INFLUENCE) CONTROL
    - POSITION AT RELEVANT FORA WITH ARGUMENTS AND EXPERIENCE
    - DISCUSS IN DETAIL IP RIGHTS AND LICENSES (**like N-C on data**)
  - **Reinforce support to Open Access/Science**
    - **CSIC Open Access Mandate (effective 1<sup>st</sup> April 2019)**
    - EXPLOIT EOSC ADVANCES (*from EOSC-Hub, EOSC-Synergy, INDIGO DataCloud, XDC, DEEP Hybrid DataCloud, Cos4Cloud*)
    - **E-INFRASTRUCTURE AND EVOLUTION OF P-P AGREEMENTS**
    - **REINFORCE TRAINING AND DISSEMINATION, NETWORKING!**
  - **Support new ideas exploiting Open Access/Data/Software**
    - **Data with privacy constraints but key for improving life quality**
    - Long term data series
    - Involve society (education and dissemination)
    - Keep in mind “AI FOR GOOD”

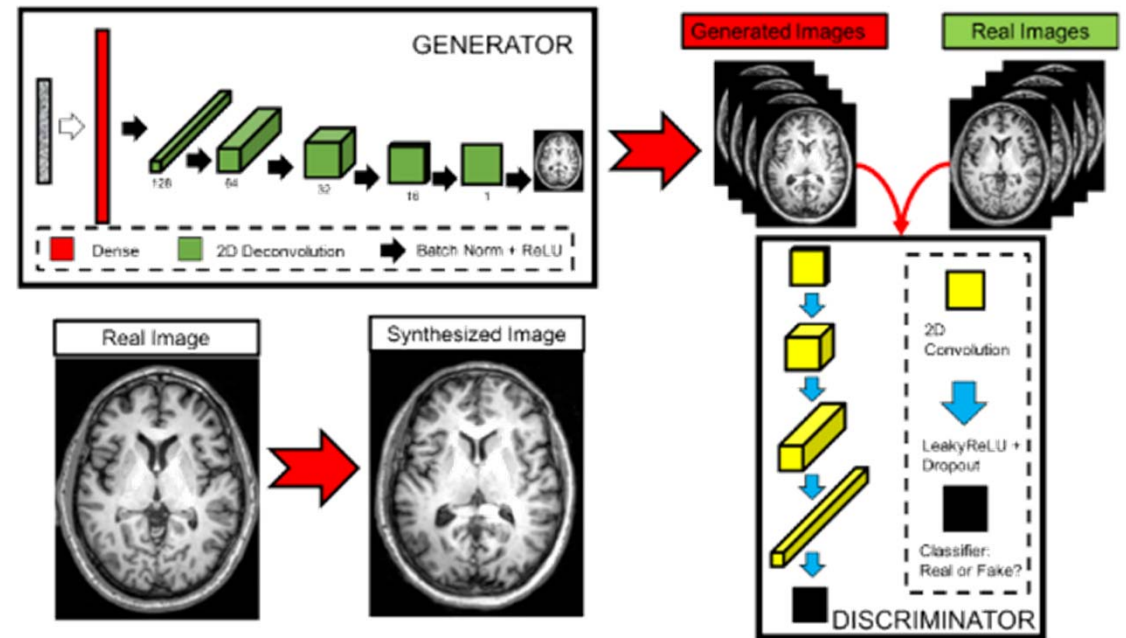




#CSIC

# What next? Data Generation

- This slide from Lara Lloret (lloret[at]ifca.unican.es)



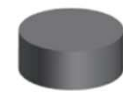
<https://thispersondoesnotexist.com/>

# What Next? Natural Language Processing (NLP)

- *This slide, also from Lara Lloret (lloret[at]ifca.unican.es)*



Electronic Medical Records



Natural Language Processing

Structured Data (Machine Interpretable)

- Classify
- Extract
- Summarize

MEDLINE Articles / Abstracts

Combining multiple sources!!!

(CT, MRI, genomics, proteomics, patient data and even hand written notes)

**A path towards Personalized Medicine**

# What next? Drug Discovery

#CSIC

www.nature.com/scientificreports

## SCIENTIFIC REPORTS

OPEN

### QSAR Classification Models for Predicting the Activity of Inhibitors of Beta-Secretase (BACE1) Associated with Alzheimer's Disease

ed: 31 January 2019  
 ed: 30 May 2019  
 ted online: 24 June 2019

From Nuria Campillo, CIB, CSIC

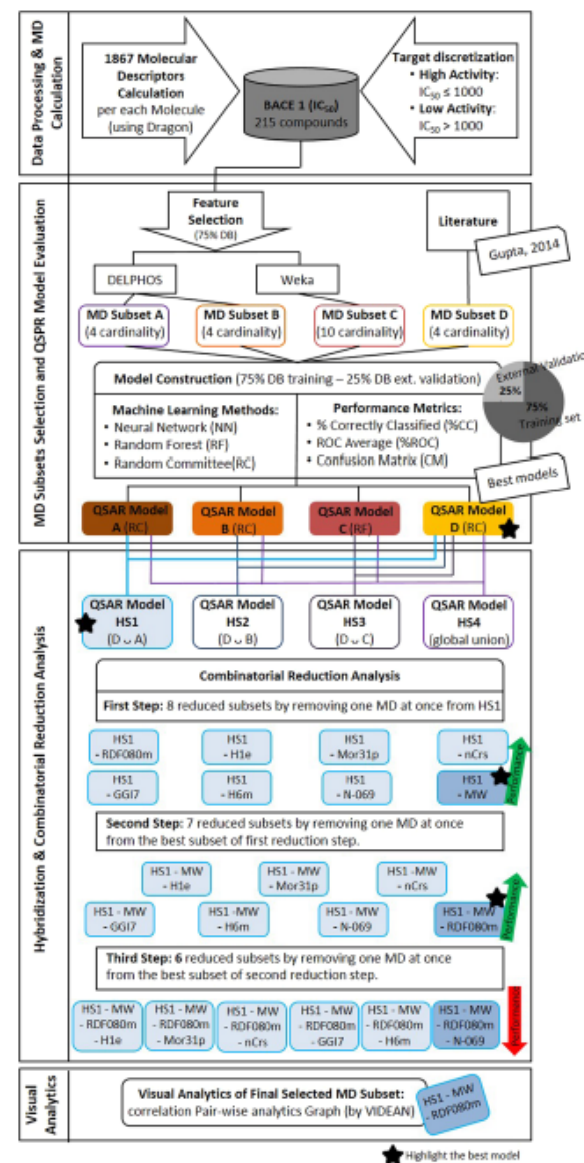


Figure 2. Graphical scheme of experiments reported for the prediction of BACE1 activity applying QSAR modelling.



# What next? CAJAL BLUEBRAIN



Javier de Felipe Oroquieta, *Instituto Cajal CSIC*  
[defelipe@cajal.csic.es](mailto:defelipe@cajal.csic.es)

#CSIC



El laboratorio Cajal de Circuitos corticales (LCCC) fue creado en 2008 como un laboratorio conjunto de investigación entre la Universidad Politécnica de Madrid (UPM, Centro de Tecnología Biomédica) y el Instituto Cajal del Consejo Superior de Investigaciones Científicas (IC-CSIC)

LCCC representa una unidad de neurociencia experimental compuesta por investigadores expertos en neurociencia (neuroanatomía y neurofisiología) del IC-CSIC y científicos computacionales e informáticos de la UPM y la Universidad Rey Juan Carlos (Madrid).

El LCCC se creó con el objetivo de combinar los estudios experimentales del cerebro con tecnologías informáticas (grupos neuroinformáticos con experiencia en estadística, herramientas informáticas y análisis de datos e imágenes), y actualmente constituye uno de los centros más avanzados que existe a nivel mundial para el estudio microanatómico del cerebro.





# A summary?

#CSIC

- Big Data hype is formally dead, long life to Big Data
- Artificial Intelligence / Data Science benefits from/requires Big Data: infrastructure, techniques, software, workflows, and datasets
- **Open Science** promotes sharing datasets and techniques
- We are only starting to explore/understand how

**Artificial Intelligence “captures” knowledge in data**

- This is key for our future: **AI assisted research**
- This trend is common to many scientific and technical area, but...  
**privacy is KEY in medicine**
- And finally our proposal: *define a research ecosystem with all stakeholders on board, and **promote data donation** to exploit all this valuable information:*

*AI for the good of our society*

***¡Únete a esta iniciativa! Gracias.***