

Gestión Forestal Sostenible basada en Ciencia de Datos



REDFORESTA 2018

Aplicación de nuevas tecnologías en el mundo forestal
Toledo | 20 al 22 de junio de 2018

Diego Rodríguez de Prado



Doctorado Conservación y Uso Sostenible de Sistemas Forestales (UVA)
Máster Gestión Forestal basada en Ciencia de Datos (UVA)
ECM Ingeniería Ambiental S.L.

Hablaremos de...

Hablaremos de...

GESTIÓN FORESTAL SOSTENIBLE BASADA EN CIENCIA DE DATOS

¿QUÉ ES LA CIENCIA DE DATOS?

Hablaremos de...

GESTIÓN FORESTAL SOSTENIBLE BASADA EN CIENCIA DE DATOS
DE LOS DATOS A LAS DECISIONES

Hablaremos de...

GESTIÓN FORESTAL SOSTENIBLE BASADA EN CIENCIA DE DATOS
APLICACIONES. DE LA TEORÍA A LA PRÁCTICA

GESTIÓN FORESTAL SOSTENIBLE BASADA EN CIENCIA DE DATOS

¿QUÉ ES LA CIENCIA DE DATOS?

Antes de nada...¿qué es un dato?

*“Es todo **material**, comúnmente **aceptado por la comunidad científica**, necesario para **validar hechos**”*

-iuFOR data-

¿Qué es la Ciencia de Datos?

*“Es un **campo interdisciplinario** cuyo fin último es **extraer conocimiento o un mejor entendimiento de los datos**, involucrando para ello una serie de **métodos científicos, procesos y sistemas**”*

-Wikipedia-



¿Qué es la Ciencia de Datos?

*“Es un **campo interdisciplinario** cuyo fin último es **extraer conocimiento o un mejor entendimiento de los datos**, involucrando para ello una serie de **métodos científicos, procesos y sistemas**”*

-Wikipedia-



DATOS



Información



Conocimiento



Decisiones

=

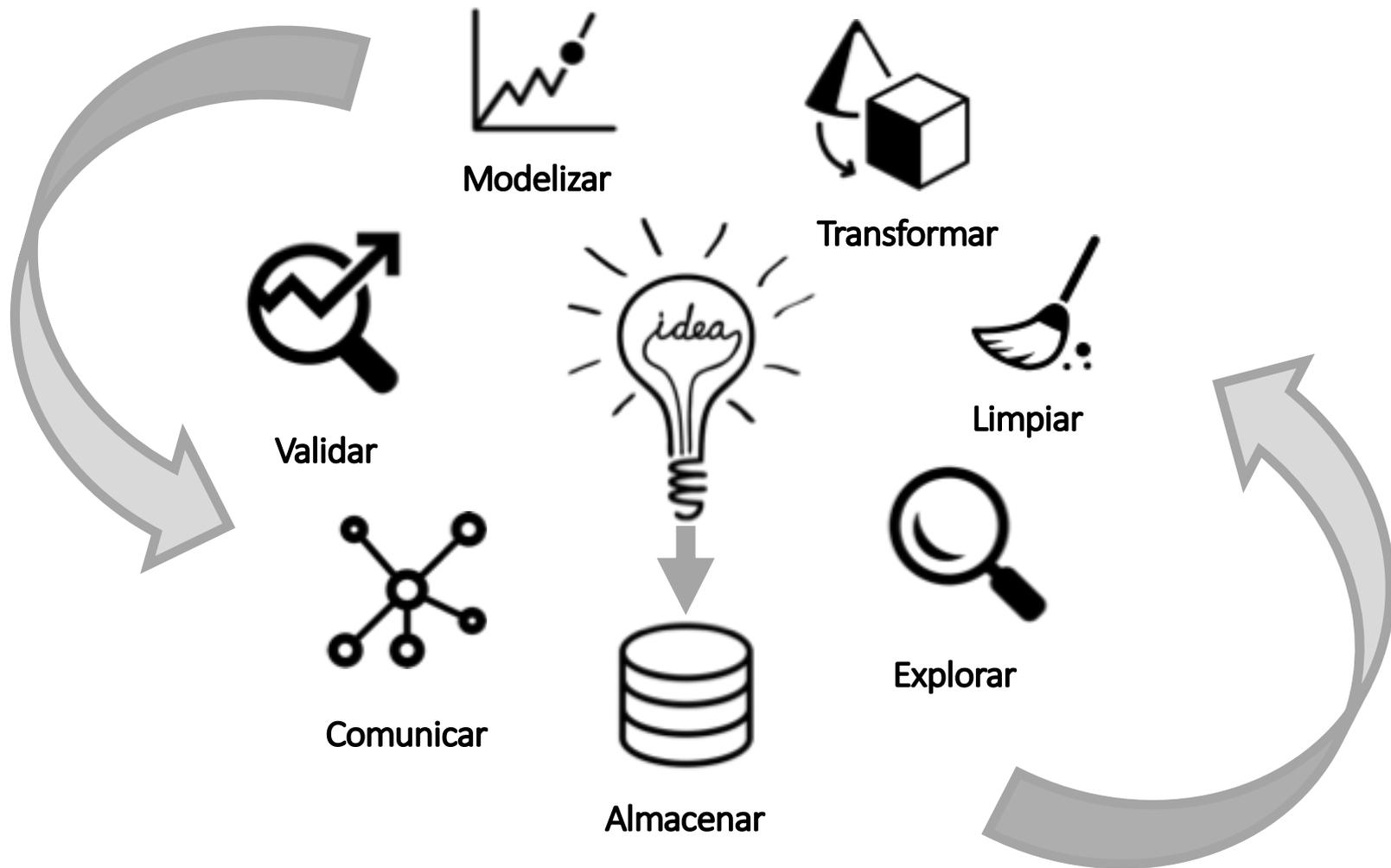
VALOR

¿Qué es la Ciencia de Datos?

*“Es un **campo interdisciplinario** cuyo fin último es **extraer conocimiento o un mejor entendimiento de los datos**, involucrando para ello una serie de **métodos científicos, procesos y sistemas**”*

-Wikipedia-

El Proceso de Ciencia de Datos



¿Qué es la Ciencia de Datos?

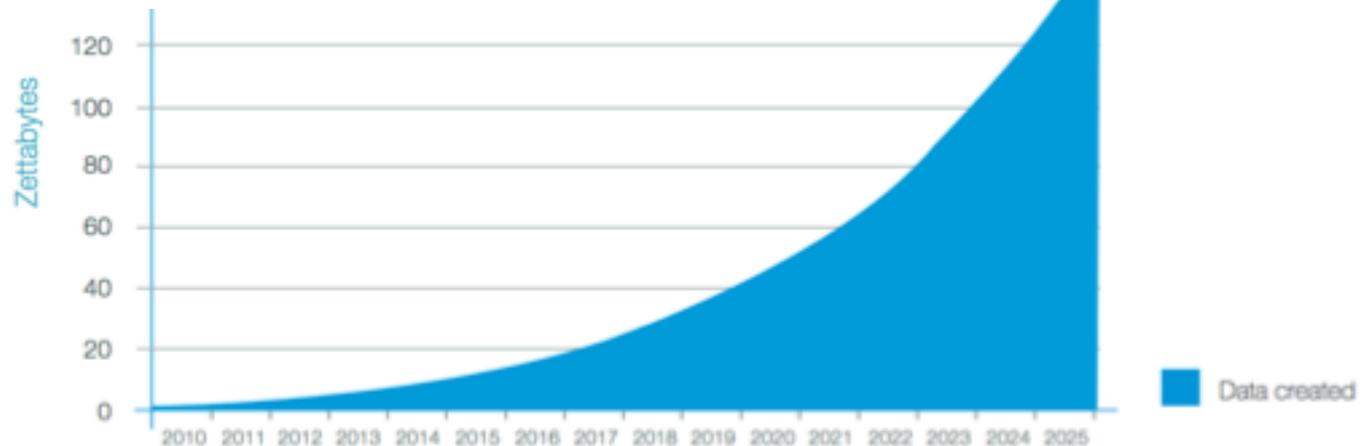
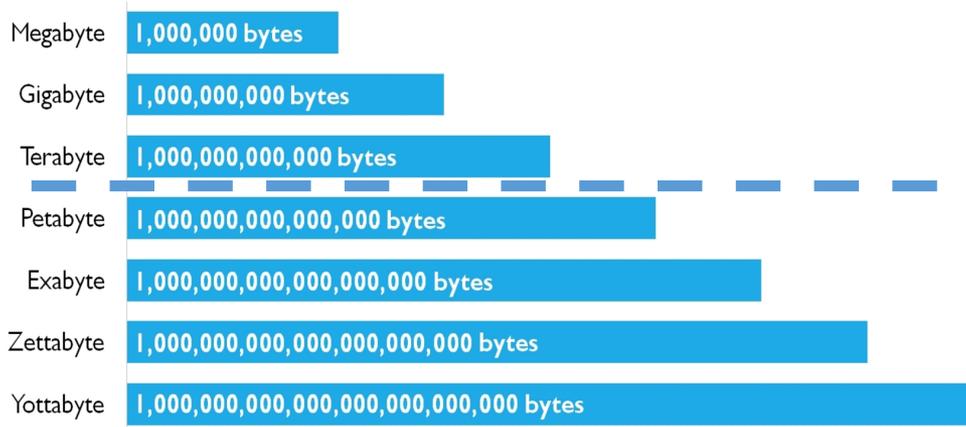
*“Es un **campo interdisciplinario** cuyo fin último es **extraer conocimiento o un mejor entendimiento de los datos**, involucrando para ello una serie de **métodos científicos, procesos y sistemas**”*

-Wikipedia-

*“Es el **cuarto paradigma de la ciencia** (empírico, teórico, computacional y ahora basado en datos). Todo lo relacionado con la ciencia está cambiando debido al impacto de la **tecnología de la información** y el **diluvio de datos**”*

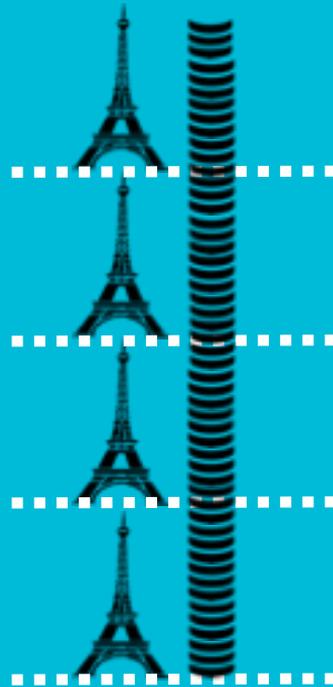
-Jim Gray -

El diluvio de datos del que habla Jim Gray



Bienvenid@s a la era del Big Data

Cada día
generamos...



90% Datos creados
entre 2011 y 2013

2.5 Exabytes
de datos

10.000.000
de discos blu-ray llenos

10%
Todos los datos
creados desde el
inicio de la historia
hasta 2011

Fuente: IBM Big Data Report 2013

Entonces, ¿qué es Big Data?



“ Gestión y análisis de **enormes volúmenes de datos** que no pueden ser tratados de manera convencional, ya **que superan los límites y capacidades de las herramientas de software habitualmente utilizadas** para la captura, gestión y procesamiento de datos “

IMC Group, 2014

Las 5 V's del Big Data



- **40.000 Exabytes** creados para 2020
- Facebook tiene **1.000 Terabytes** de almacenamiento

- **90%** de los datos generados están desestructurados
- **60%** es video

Cada minuto:

- **21.000.000** whatsapps
- **350.000** tweets

- **1/3** directivos no confía en los datos para tomar decisiones

Big Data Landscape 2016

Infrastructure

Hadoop On-Premise
cloudera, Hortonworks, MAPR, Pivotal, IBM InfoSphere, splice, bluedata, Jethro

Hadoop in the Cloud
Amazon Microsoft Azure, Google Cloud Platform, IBM InfoSphere, CAZENA, otisclose, Dcode, aplyntly

Spark
databricks, GridGain, TACHYON NEXUS

Cluster Services
amazon, docker, MESOSPHERE, CoreOS, prepdata, StockIQ

Analytics

Analyst Platforms
Palantir, AYASDI, Quid, enigma, Digital Reasoning, COMBIMONITOR

Analytics Platforms
Microsoft, guuvus, Datasense, InterJane

Data Science Platforms
Consistent relevant, CONTINUUM, DataRobot, Alpine, HODGE, SALTTRU, DOMINO, sense, THAT, ALGORITHMIA

Visualization
tableau, Google Data Studio, Rookzbi, SCORIX, Qlik, CHARTIO

Applications

Sales & Marketing
RADIUS, Gainsight, bloomreach, Zeta, blueyonder, Livefyre, Lattice, Persado, infer, sense, AVISO, ACTIONIQ, QUANTIPRA, ENGADIG

Customer Service
MEDALLIA, ATTEUNITY, STERILIA Service, NGDATA, DigitalGaming, epurati, face machines

Human Capital
gild, Connect for, entelo, entelo hiQ

Legal
RAVEL, LUDICUS, Everlaw, Brevia, HUBERINTOX

NoSQL Databases
amazon DynamoDB, Google Cloud Platform, Microsoft Azure, mongoDB, KEROSPIKE, Couchbase, Sequoia, redislabs, influxdata

NewSQL Databases
SAP, Clustring, Pivotal, paradigm4, memsql, NUOQB, MariaDB, VOLTDB, citusdata, loopdb, Trafalgar, Cockroach Labs

BI Platforms
Power BI, amazon, Domo, Wave Analytics, GoodData, platform, atscale

Statistical Computing
sas, SPSS, MATLAB

Log Analytics
splunk, sumologic, loggly

Social Analytics
NETBASE, DATASFT, treckr, bitly, synthetix, simpleReach

Graph Databases
oneoj, OrientDB, InfoGraph

MPP Databases
TERADATA, VERTECA, HIVEPERA, Kognitio, dremio

Cloud EDW
amazon, Google Cloud Platform, Microsoft Azure, Pivotal, snowflake, AWS Redshift, IBM InfoSphere

Data Transformation
alteryx, TRIFACTA, tamr, Pentaho, StreamSets, D-7 Alation

Data Integration
informatica, MuleSoft, snaplogic, FedratiData

Real-Time
amazon, WITSOBREETS, confluent, DataArtisans

Machine Learning
IBM Watson, Cortana, VIV, DataRobot, H2O, Skytree, DataFusion, Predictio3D, givtrak

Speech & NLP
NarrativeScience, apical, Nuance, semanticworks, IBM Bluebird, iSIBOR, vscope

Horizontal AI
IBM Watson, Cortana, VIV, DataRobot, H2O, Skytree, DataFusion, Predictio3D, givtrak

Management / Monitoring
New Relic, AFFDYNAMICS, amazon, octifio, splunk, Yricono

Security
TANUUM, Burmio, CODE42, DataGravity, OpenCloud, VECTRA, Splunk, Bluefish

Storage
amazon, Google Cloud Platform, Microsoft Azure, panasonic, nentle storage, Quimulo

App Dev
apigee, CRSK, Typesafe, CONCURRENT

Crowd-sourcing
amazon, WorkFusion

Search
amazon, ORACLE, ELASTICSEARCH, Lucidworks, MAANA, swiftype, Algolia, sinuous

Data Services
IBM OPERA, kaggle, DataKind

For Business Analysts
Original Logic, ClearStory, CIRRO, Import IO

SMB / Commerce
Google Analytics, Amazon, RJMetrics, BLUECORE, sumal, granify, retersee, custora

Publisher Tools
Outbrain, mixpanel, Chartbeat, yieldbot, Yieldmo

Govt/ Regulation
Socrata, OPENGOV, FiscalNote, enigma, PROZOR, mark43, OpenDataSoft

Finance
affirm, LendingClub, OnDeck, Kreditech, Kabbage, INSIKT, Lenddo, KENSHC, AIDYA, ISENTIUM, Quantoplan, sentient

Cross-Infrastructure/Analytics
amazon, Google, Microsoft, IBM, SAP, SAS, hp, VMware, talend, TIBCO, TERADATA, ORACLE, NetApp

Framework
Hadoop, YARN, Spark, HESOS, TEZ, Flink, CDAP

Query / Data Flow
SLAMDATA, DRILL, Amazon Cloud EMR

Data Access
SCOUT240, HBASE, mongoDB, cassandra, riak, COUCHDB, PSYCOPE, ORACLE, DRYICE

Coordination
talend, Apache Zookeeper, Apache Ambari

Real-Time
STORM, Spark, Flink, TACHYON, druid

Stat Tools
R, Scala, SciPy

Machine Learning
mllib, Aerospike, Caffe, NINGA, CNTK, FeatureFu, DIMSUM, DLAI

Search
elasticsearch, Solr

Security
Apache Ranger, Visualization

Health
Apple, JAWBONE, GARMIN, practicefusion, fitbit, Withings, VALIDIC, netatmo, kinsa, Human API

Financial & Economic Data
Bloomberg, Dow Jones, YODLEE, PREMIERE, ESP CAPITAL IQ, quora, xignite, CB Insights, marketmark, Gestrize, PLAB

Air / Space / Sea
PLANET LAB, spire, AIRWARE, DroneDeploy

Location/People/Entities
GARMIN, foursquare, InsideView, esri, STREETLINE, cambo, factual, PlaceIQ, Crosson Maps, placemeter, BASIS, Sense

Other
qualtrics, panjiva, DATA.GOV, GA, DataCamp, INSIGHT, DataEline, METI, The Data Incubator

L@s científic@ de datos



"Persona que sabe más de estadística que cualquier programador/a y que a la vez sabe más de programación que cualquier estadístico/a"

- Josh Wills -

Plan de Manejo de Datos (DMP)

Es el documento que describe de forma específica **como serán tratados los datos** generados en un proyecto.

Los datos deben ser planificados porque:

- Ayuda a darle **organización**, de modo que todos los miembros del equipo investigador sepan la ubicación exacta del dato y todos sepan las condiciones de recolecta o como se generó dicho dato
- Contribuye a generar una investigación mucho más **participativa**, ya que ayuda a validar resultados de la propia investigación o demás investigaciones
- Mejora la **visibilidad** de la investigación en el mundo científico, dando la posibilidad a los datos de ser utilizados de otras formas que el investigador o grupo investigador no habían imaginado
- Además, es **obligatorio**

GESTIÓN FORESTAL SOSTENIBLE BASADA EN CIENCIA DE DATOS
DE LOS DATOS A LAS DECISIONES

De los Datos a las Decisiones en la Gestión Forestal Sostenible



Datos

*Inventario Forestal
Imágenes satélite
LiDAR*



Información

*Forestal
Economía
Ecología*



Conocimiento

*Crecimiento
Productividad
Servicios Ecosistémicos*



Decisiones

*¿Cuántos pies cortar?
¿Ordenación de masa pura o mixta?*

Y los forestales...¿con qué datos trabajamos?

REDFORESTA 2018
Toledo | 20 al 22 de junio de 2018



Y los forestales...¿con qué datos trabajamos?



“ Proyecto encaminado a obtener el máximo de información posible sobre la situación, régimen de propiedad y protección, naturaleza, estado legal, probable evolución y capacidad productora de todo tipo de bienes de los montes españoles “

- Menor disponibilidad que otros sectores (*mayor dificultad, menores recursos*)
- Privacidad de los datos

*Ministerio de Agricultura y Pesca,
Alimentación y Medio Ambiente*



2007-
2017

Y los forestales...¿con qué datos trabajamos?



Drones, LiDAR y Satélites, de la escasez a la abundancia de datos

- Menor disponibilidad que otros sectores (*mayor dificultad, menores recursos*)
- Privacidad de los datos



Y los forestales...¿con qué datos trabajamos?



Estandarización e interrelación de Bases de Datos de diferente naturaleza
Datos **abiertos** y **enlazados**

- Diferentes escalas de trabajo, temáticas, formatos y fuentes
- Sistema vivo, relaciones entre variables

IFN
Inventario Forestal Nacional



Y los forestales...¿con qué datos trabajamos?

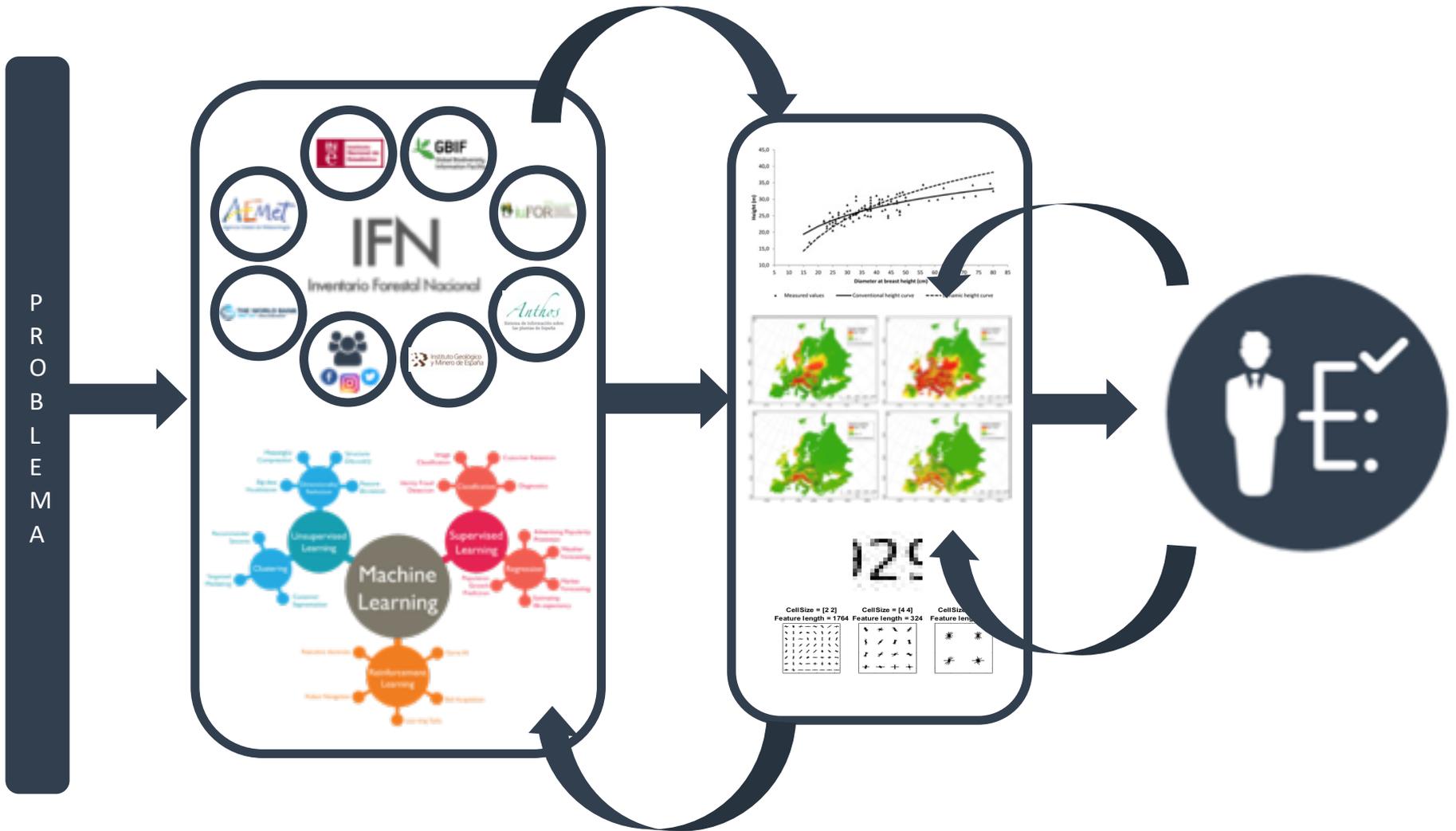


- Datos normalmente estáticos, temporales
- Hacia una toma de datos dinámica, en tiempo real



- Datos sujetos a diferentes tipos de errores
- Mismo dato, diferentes fuentes

De los Datos a las Decisiones en la Gestión Forestal Sostenible



De los Datos a las Decisiones en la Gestión Forestal Sostenible



De la Inferencia Estadística Clásica al Aprendizaje Automático

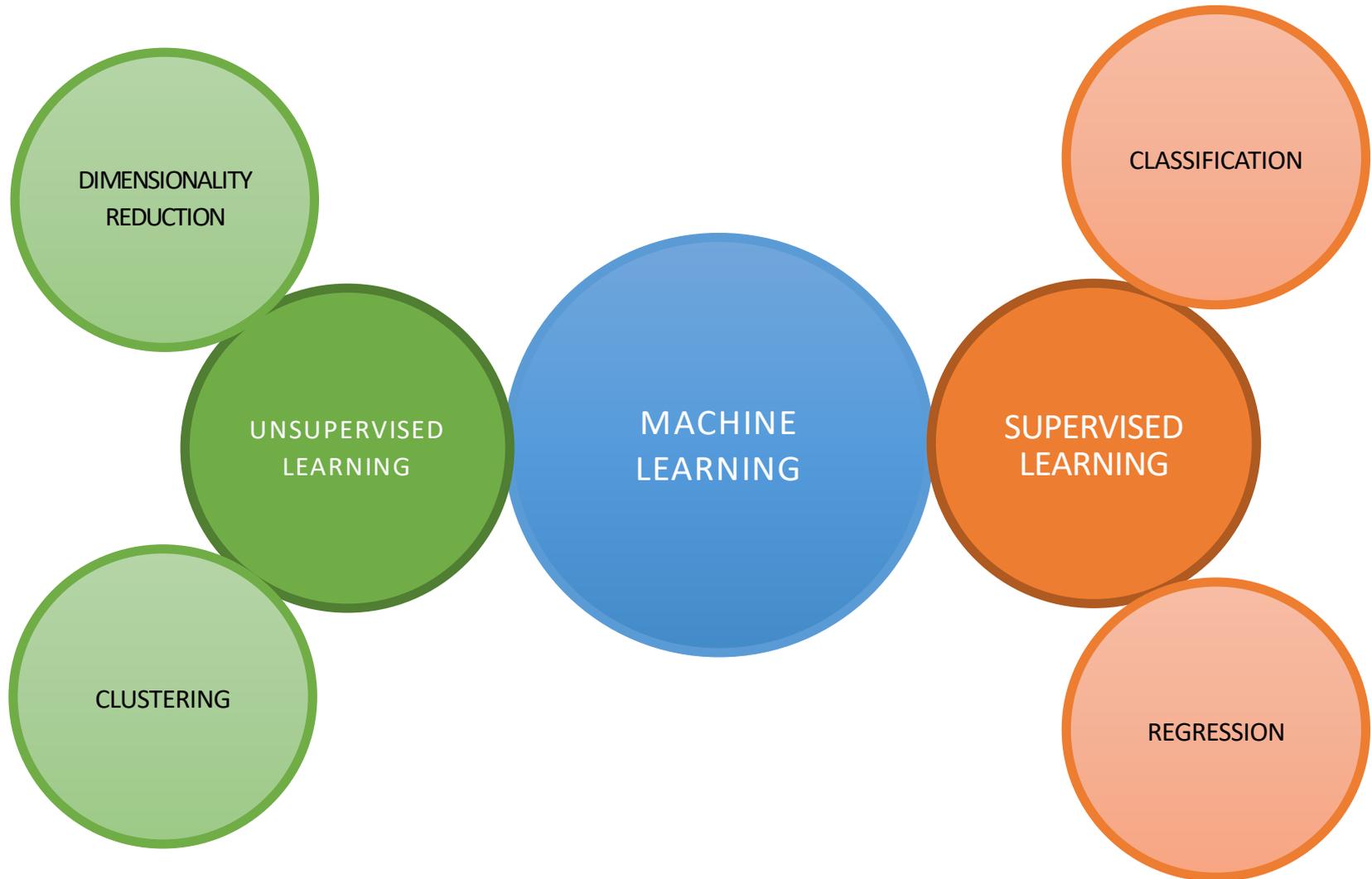
Inferencia Estadística Clásica

Conjunto de métodos y técnicas que permiten inducir, a partir de la **información empírica** proporcionada por una **muestra**, cual es el comportamiento de una determinada población con un riesgo de error medible en términos de probabilidad

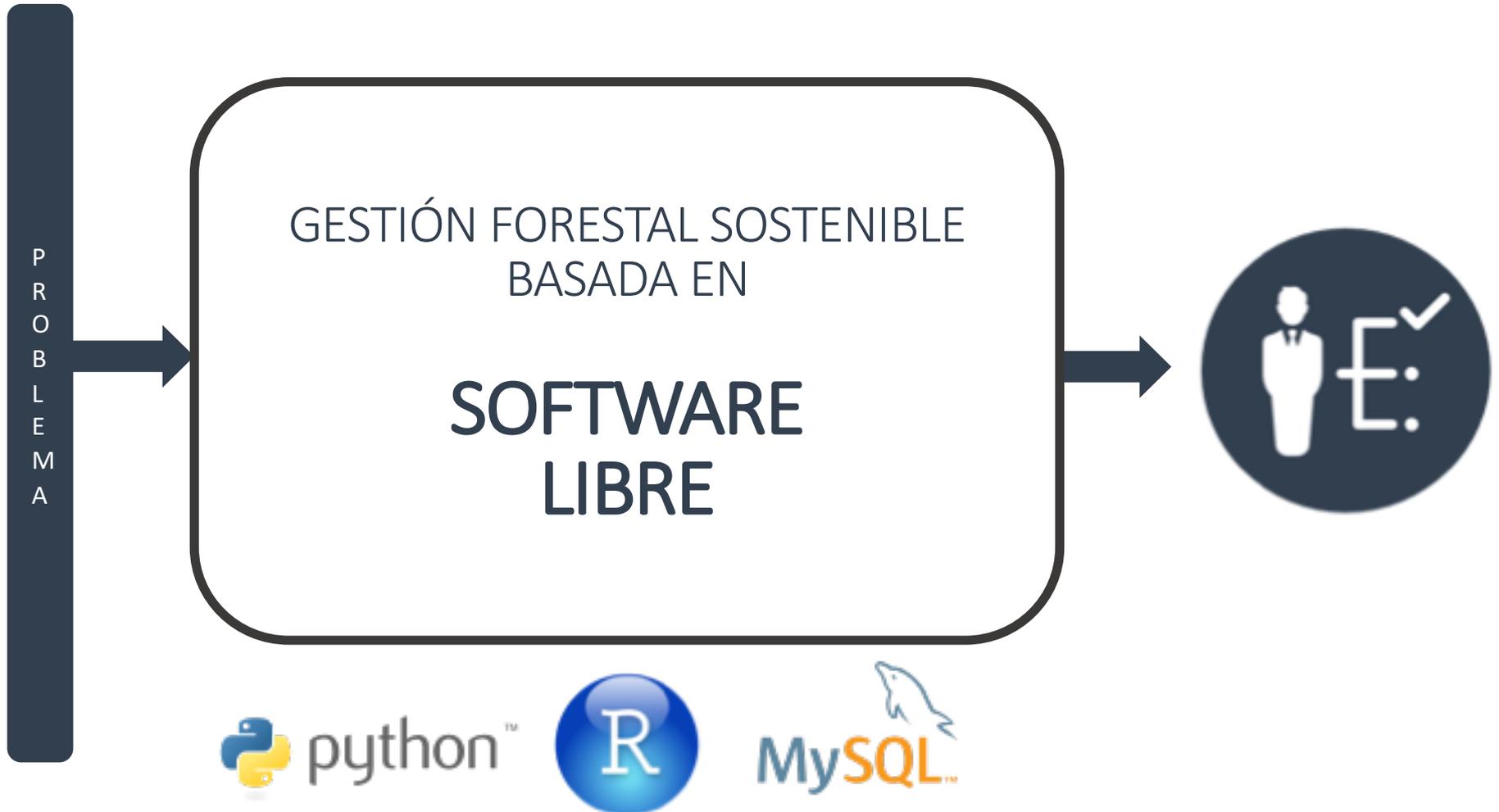
Aprendizaje Automático | MACHINE LEARNING

Tipo de Inteligencia Artificial que proporciona a las computadoras la capacidad de aprender, sin ser programadas explícitamente. El aprendizaje automático se centra en el desarrollo de programas informáticos que pueden cambiar cuando se exponen a **nuevos datos**.

De la Inferencia Estadística Clásica al Aprendizaje Automático



De los Datos a las Decisiones en la Gestión Forestal Sostenible



De los Datos a las Decisiones en la Gestión Forestal Sostenible



GESTIÓN FORESTAL SOSTENIBLE BASADA EN CIENCIA DE DATOS APLICACIONES. DE LA TEORÍA A LA PRÁCTICA

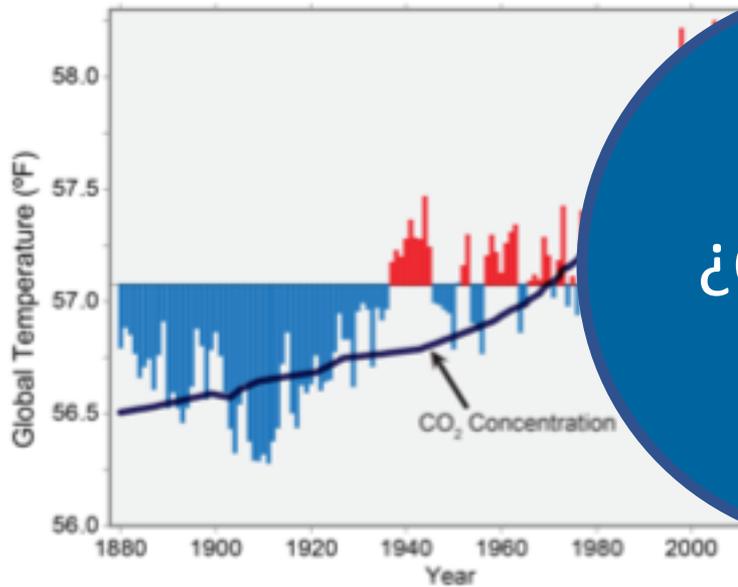




QUANTC®

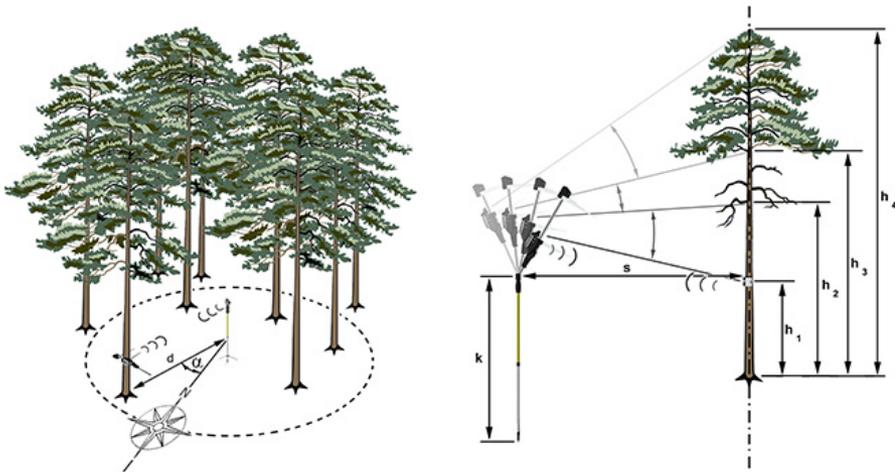
Una Herramienta para la Cuantificación de Carbono en Sistemas Forestales
Herrero de Aza.C; De la Parra.B; Rodriguez de Prado.D. 2017

Global Temperature and Carbon Dioxide



¿CUÁNTO?

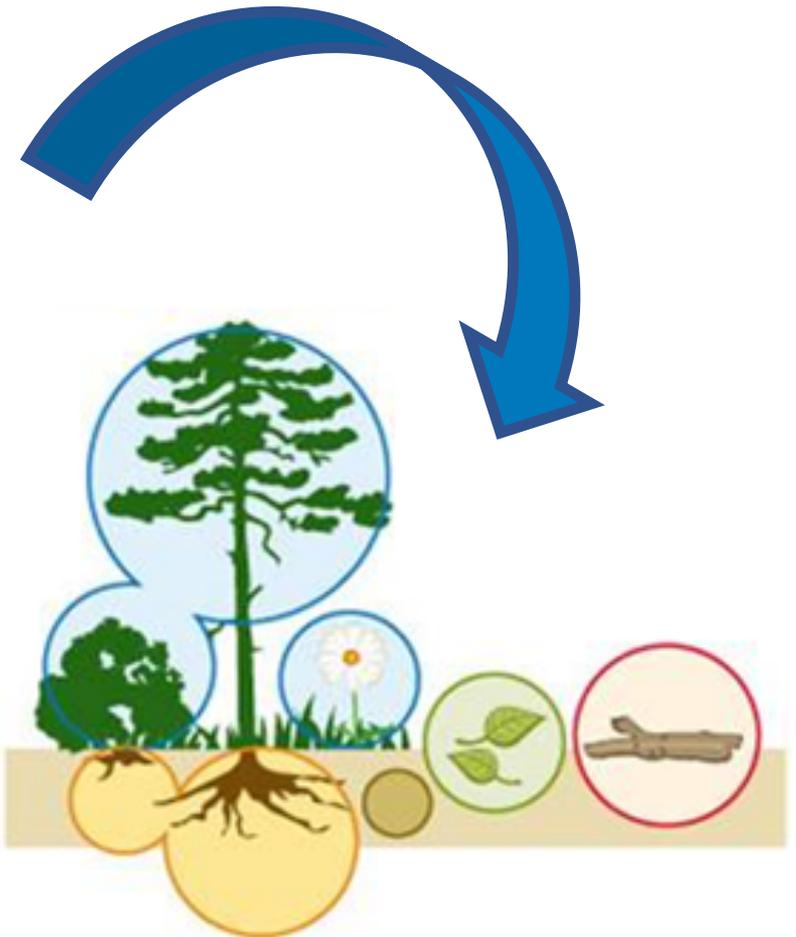
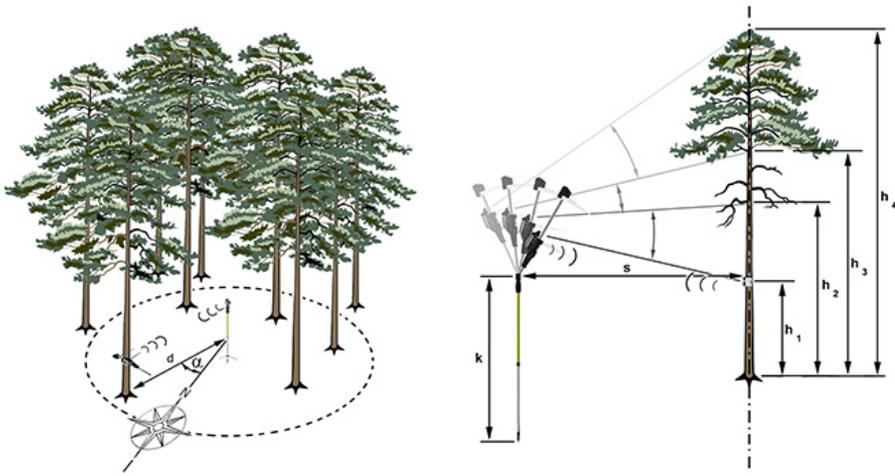




Ecuaciones de Biomasa

+

Inventarios Forestales



Biomasa

Carbono

CO₂



Potencialidad
Biomasa

Huella de
Carbono

Certificación
Emisiones

Modelización
aplicada



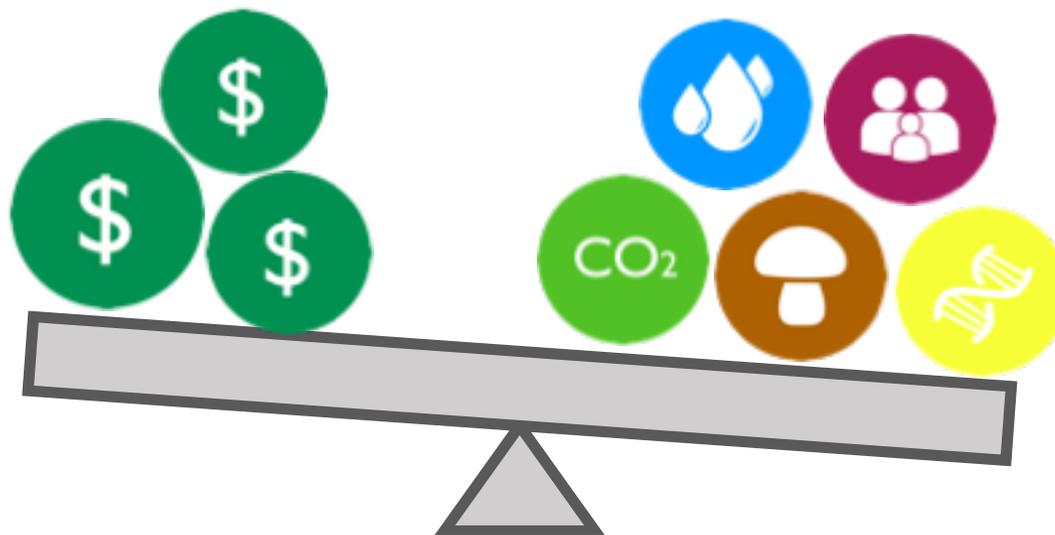
Smartelo®

Una Herramienta para la Gestión, Análisis y Presentación de Datos en Parcelas Forestales
Rodríguez de Prado.D; Bravo.F; Ordoñez.C. 2016

Cambio de paradigma en la Gestión Forestal

REDFORESTA 2018
Toledo | 20 al 22 de junio de 2018

Gestión Forestal
Tradicional
(Productividad)



Gestión Forestal
Sostenible
(Multifuncionalidad)

Marteloscope, una nueva herramienta para la Gestión Forestal Sostenible

Parcelas forestales permanentes

Con un área estándar de **1 ha**
divididas en **16** sub-parcelas o cuadrantes
donde todos los pies son medidos,
identificados y geolocalizados.

Max Bruciamacchie (AgroParisTech – ENGREF, Nancy)

Marteloscopes, parcelas forestales creadas para...



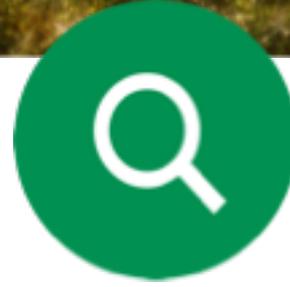
FORMACIÓN



TRANSFERENCIA



TRABAJOS
TÉCNICOS



I+D+i



NETWORKING

...y también para jugar con datos

smartevo

¿Qué?

¿Por qué?

Aplicación open-access desarrollada en Microsoft Excel y VBA (*Visual Basic for Applications*), diseñada para gestionar, analizar y presentar datos forestales, ecológicos y económicos en marteloscopes

- **Apoyar** actividad señalamiento forestal
- **Potenciar** el uso de nuevas tecnologías y ciencia de datos en gestión forestal sostenible
- **Concienciar** población y stakeholders



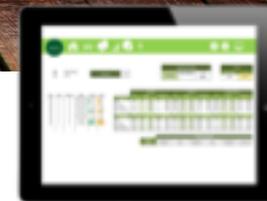
Smartelo PC | Versión de escritorio

Análisis forestal, ecológico y económico del marteloscope

Output numérico y gráfico

Comparación de señalamientos entre diferentes gestores/técnicos/equipos

Análisis de biomasa y productos de transformación de la madera



Smartelo Portable | Versión Mobile-friendly

Seguimiento de la tarea de señalamiento in situ y en tiempo real

Determinación de diferentes objetivos selvícolas centrados en multifuncionalidad

Información numérica y gráfica de la actividad de señalamiento forestal

Cómo podemos utilizar Smartelo para fines didácticos?



Rodríguez de Prado, D. In Bravo Oviedo F.,dir, Ordóñez Alonso C.,dir.(Eds.), 2016. *Smartelo: Una herramienta para el cálculo, gestión y presentación de datos en aulas de señalamiento forestal*



Objetivo:

Potenciar la complementariedad entre Smartelo y nuevas herramientas basadas en Ciencia de Datos

REDFORESTA 2018
Toledo | 20 al 22 de junio de 2018

Smartevo
The Smart Forest Management





SIMANFOR®

Sistema de Apoyo para la Simulación de Alternativas de Manejo Forestal Sostenible
Bravo.F; Ordoñez.C. 2016

Para finalizar...



DATOS



Información



Conocimiento



Decisiones
=
VALOR

El Proceso de Ciencia de Datos



Modelizar



Transformar



Validar



Limpiar



Comunicar

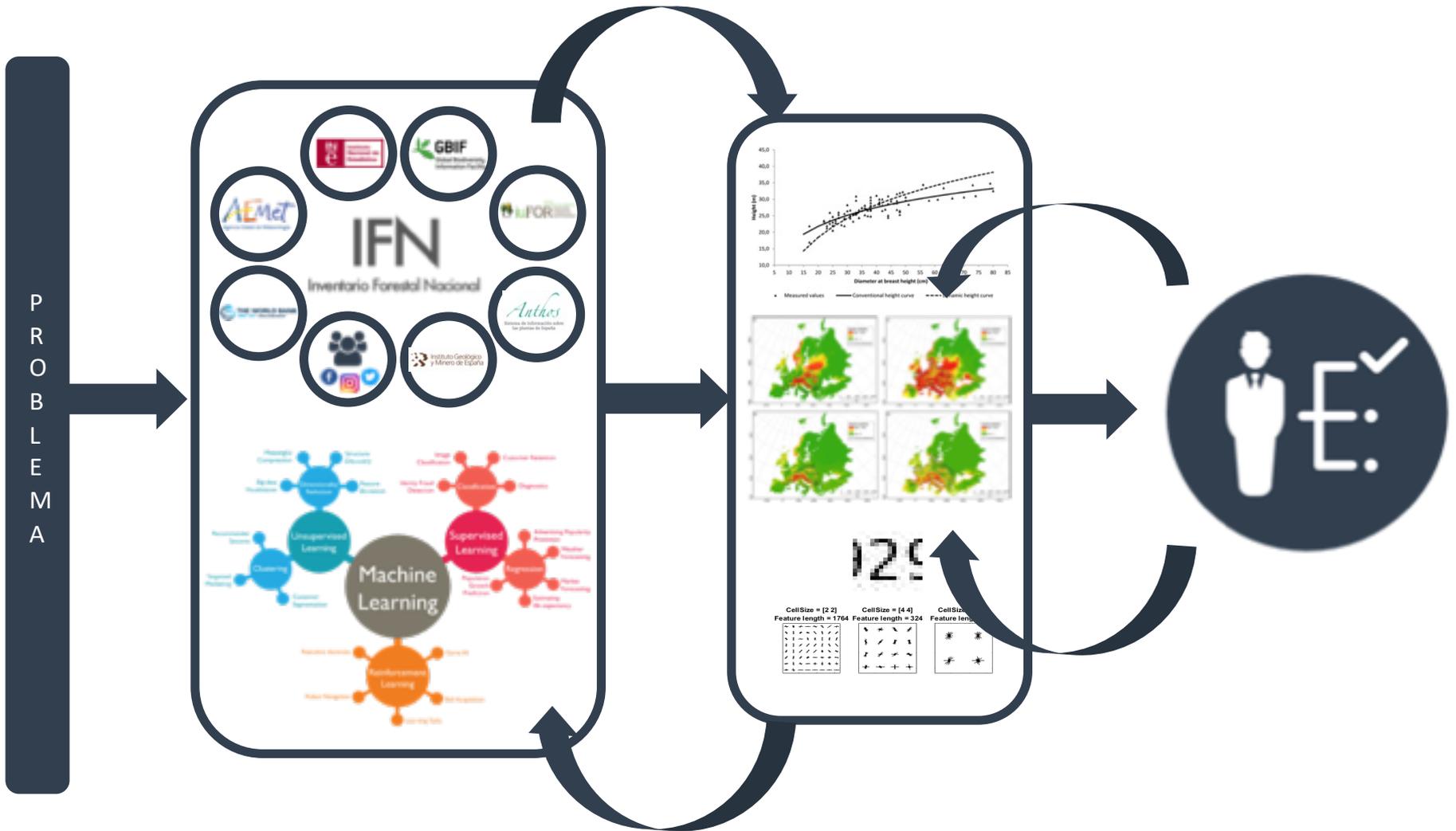


Explorar

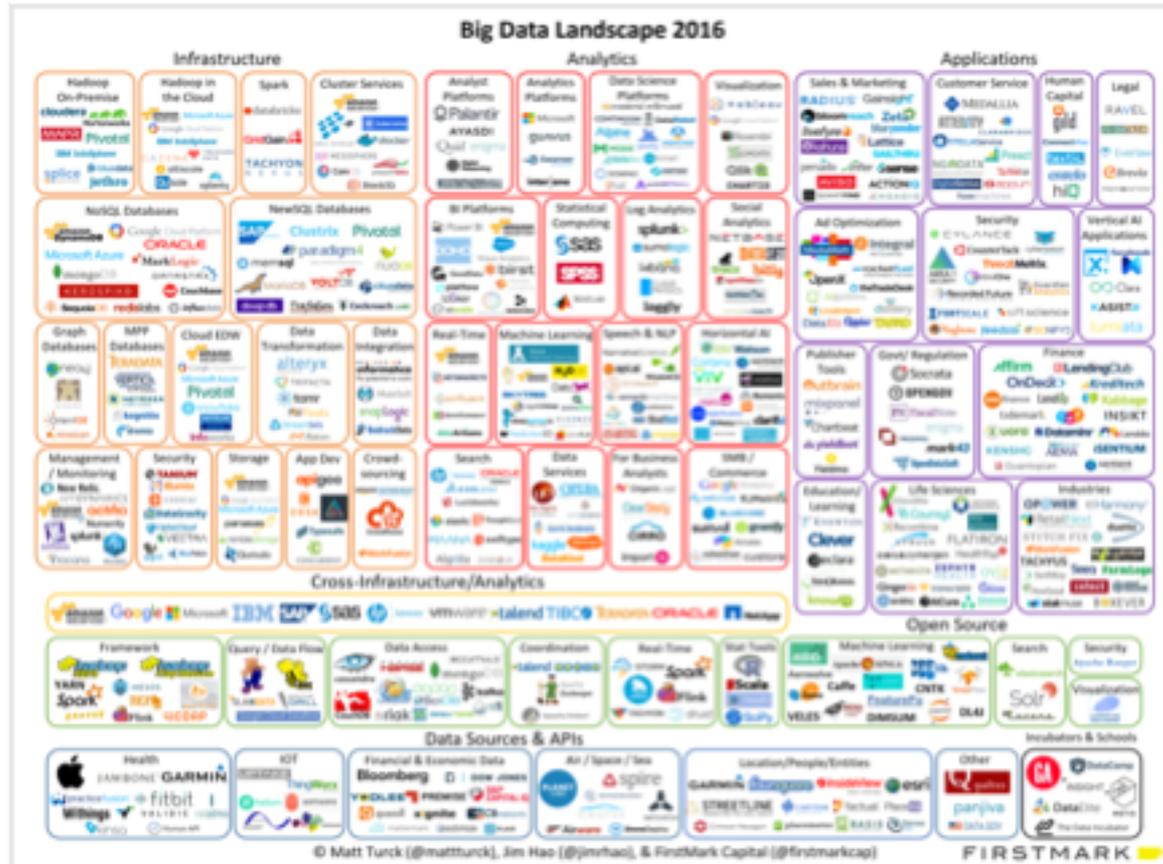


Almacenar

Para finalizar...



Para finalizar...



MUCHAS GRACIAS :)



REDFORESTA 2018

Aplicación de nuevas tecnologías en el mundo forestal
Toledo | 20 al 22 de junio de 2018

Diego Rodríguez de Prado

diego.rodriguezdeprado@gmail.com
@dataforester



Gestión Forestal Sostenible basada en Ciencia de Datos



REDFORESTA 2018

*Aplicación de nuevas tecnologías en el mundo forestal
Toledo | 20 al 22 de junio de 2018*

Diego Rodríguez de Prado

diego.rodriguezdeprado@gmail.com

@dataforester

