

Introducción a la ciencia de datos en agricultura

Hugo Andrés Dorado

Científico de datos

hugo.doradob@gmail.com

Contexto Internacional

CUARTA REVOLUCIÓN INDUSTRIAL ▶

Intensiva en conocimiento, desarrollo tecnológico, innovación y explotación de los datos



1.^a

1784

Producción mecánica
y energía de vapor



2.^a

1870

Producción en masa
de energía eléctrica



3.^a

1969

Electrónica y TI



4.^a

HOY

Sistemas físicos cibernéticos
(Big Data, IoT, Smart Cities,
inteligencia artificial)

◀ **DATIFICACIÓN:**
transformación del
mundo en datos
procesables y
cuantificables

~ Crecimiento de datos en el mundo (billones de gigabytes) ~

1,2
billones
2010

16,1
billones
2016

163
billones
2025

2017



1,9 billones
Búsquedas



84,4 billones
Correos enviados



241.000 millones
Tweets



2,2 billones
Videos vistos



25.000 millones
Fotos subidas



32,2 millones
Páginas hackeadas



89.000 millones
Llamadas

Cambios tecnológicos en la agricultura

Agricultura tradicional



Agricultura digital, colectando y diseminando datos a gran escala



Colectando datos a gran escala (IoT)

Soil moisture sensors



Connected pest traps



Connected machinery



RFID chips



Drone imagery



Smaller cheaper weather stations



New sensor technologies



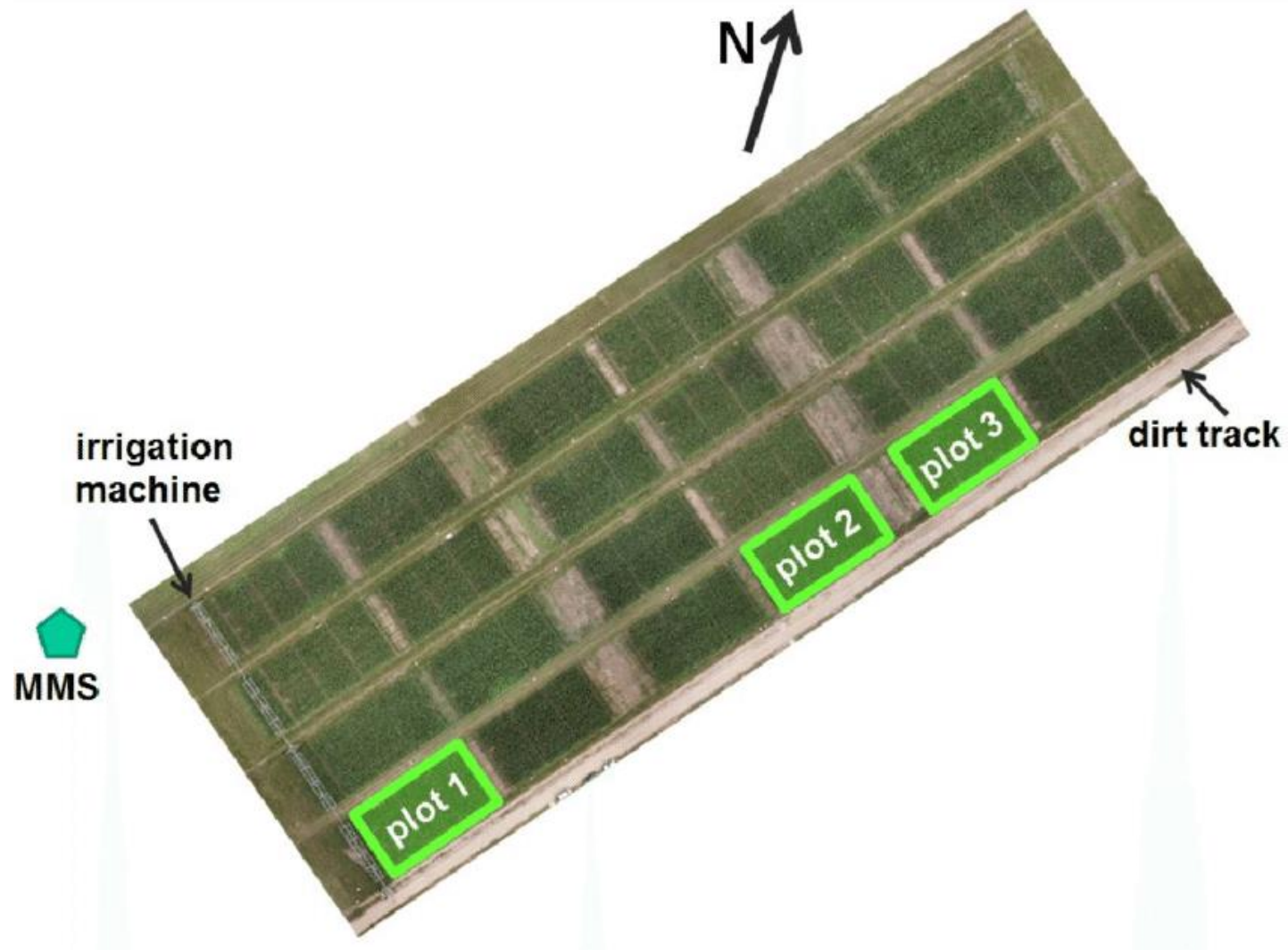
Agricultura guiada por los datos.



Tipos de estudios en agricultura

- Experimentales.
- Observacionales.

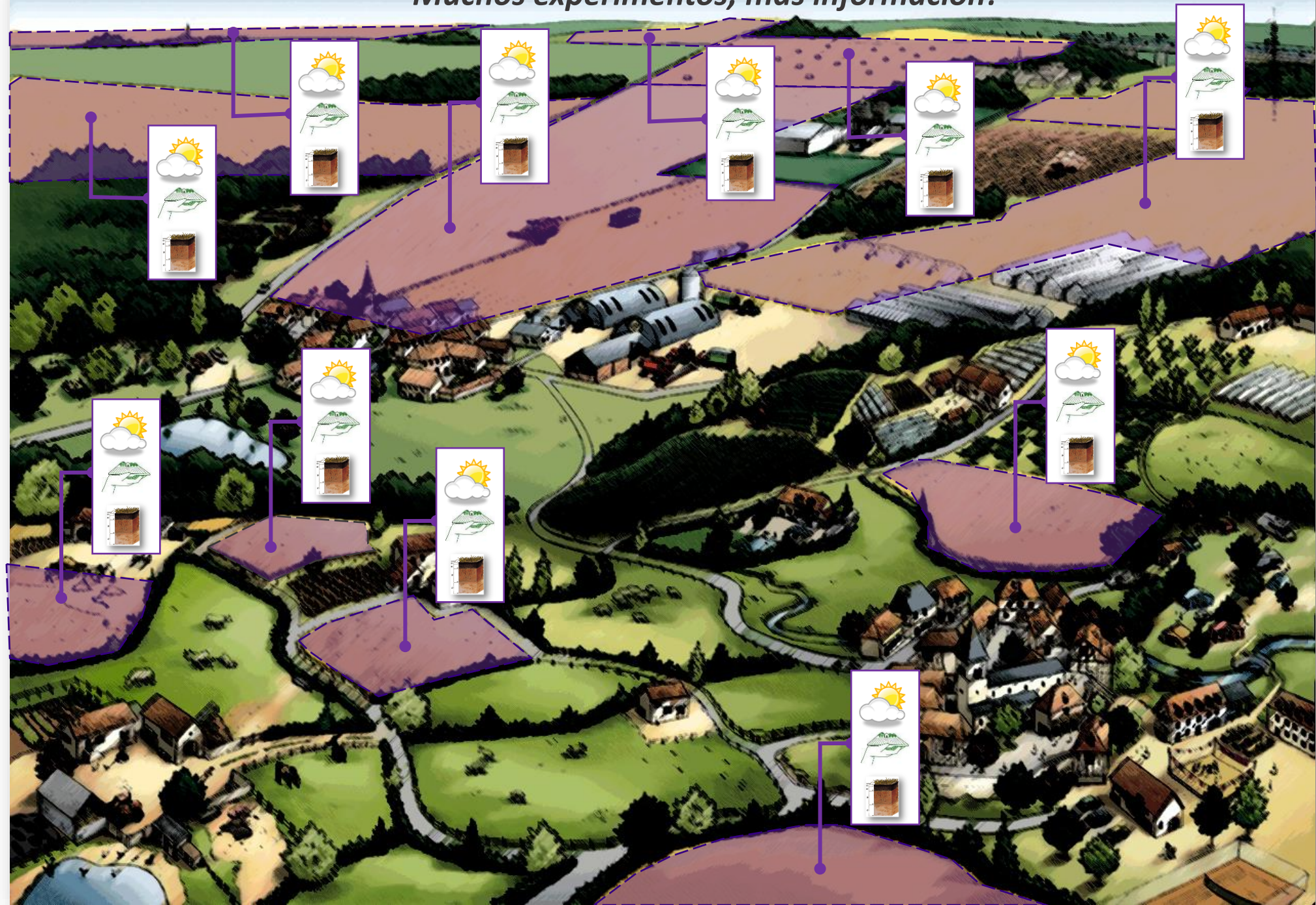
Estudios experimentales



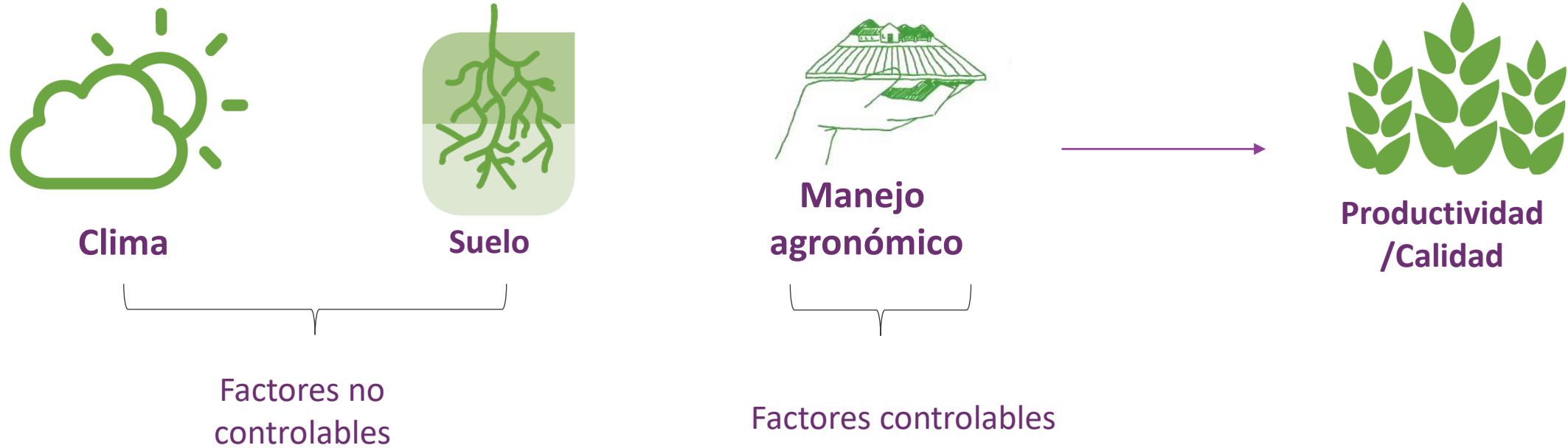
A gran escala, estudios observacionales



Muchos experimentos, más información!



Agricultura específica por sitio



Definición:

“El arte de tomar de decisiones en agricultura basados en las condiciones biofísicas (Clima y suelo) del cultivo”

Unidad de manejo

Edad de plantas

Tipo de suelo



Tipos de manejo

Variedades

Ciencia de los datos

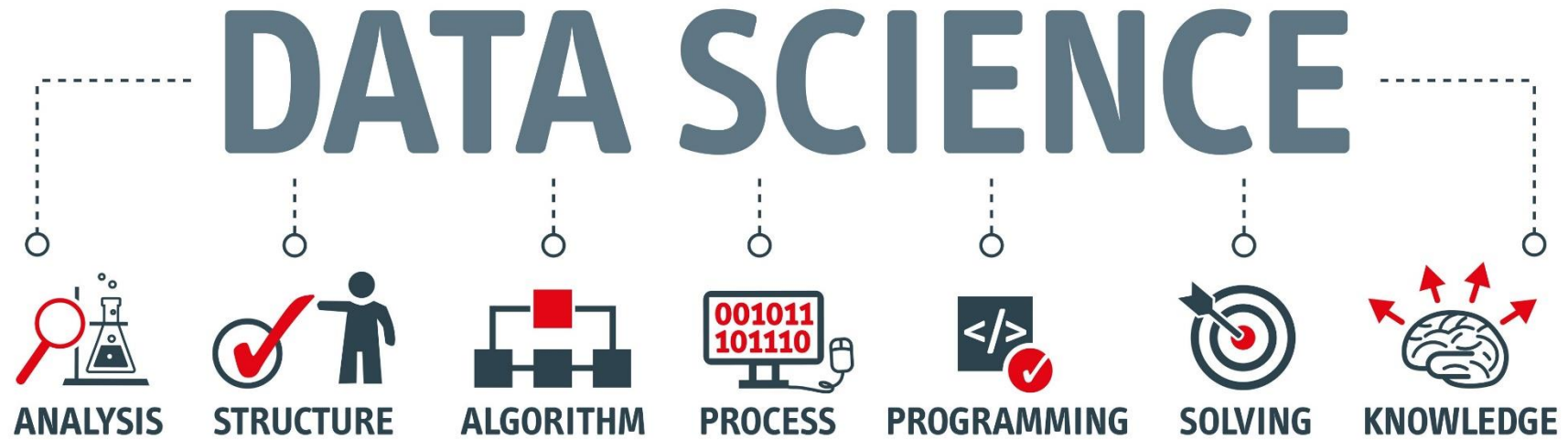
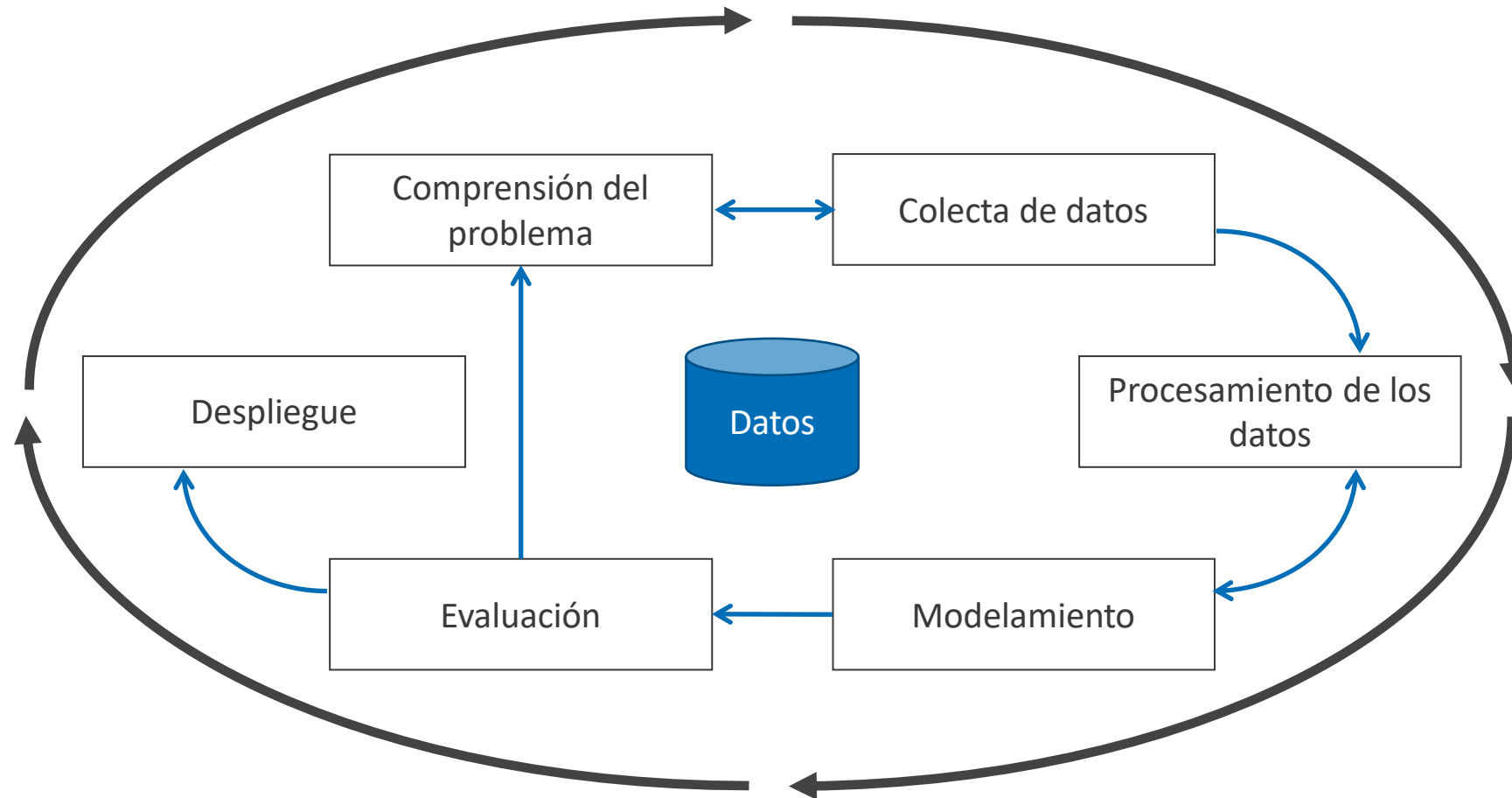


Image by shutterstock from Datanami

Flujo de trabajo en análisis de información



Basado en

CRISP-DM
CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING

<https://data.sngular.com/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science>

Fuentes de información

Información primaria

Registros de cosecha.



Plataformas webs para colecta y reporte.



Aplicacion es móviles.



Drones



Información Secundaria

Estaciones meteorológicas



Percepción remota.



Decisiones impulsadas por datos



Datos históricos.

Clasificar las siguientes fuentes de información como primarias o secundarias.

- Los registros de campo del señor José que ha colectado por 5 años y serán analizados en conjunto con datos de otros productores.
- Los pluviómetros del Senamhi en la zona Norte de país, que serán utilizados para zonificación agroecológica.
- Los datos colectados por la cosechadora de arroz, utilizada para optimizar el sistema productivo.
- Las coordenadas geográficas de los lotes de una finca para mapear ArcGis.

Colectando datos a gran escala (Iniciativas de datos abiertos)



Worldclim

Chirps

Nasa Power

SoilGrid

Instituciones públicas

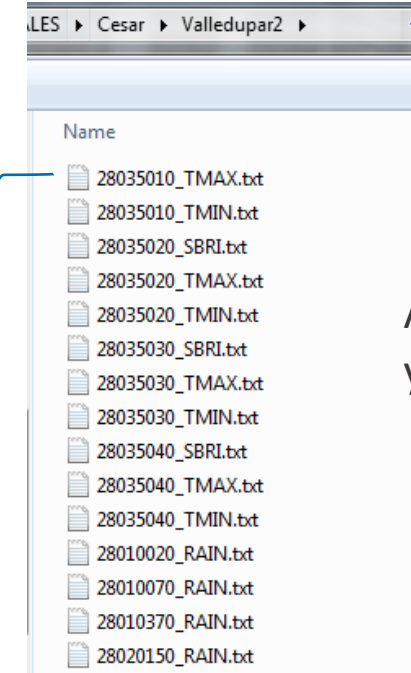


Procesamiento y limpieza de la información

Distintos formatos de fecha y caracteres

A	B	C	D	E	F	G	H
AuxVar	Y	Lote	FechaSiembra	FechaCosecha	Tecnico	Area	Cod. Act.
CHARCO-10_1_F	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	9 1_Fitosani
CHARCO-10_2_	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	12 2_Fitosani
CHARCO-10_Co	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	45 Control_G
CHARCO-10_Co	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	8 Control_P
CHARCO-10_Co	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	36 Control_P
CHARCO-10_De	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	19 Desinfecci
CHARCO-10_Fe	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	27 Fertilizaci
CHARCO-10_Fe	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	27 Fertilizaci
CHARCO-10_Fo	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	2	31 Foqueo_F
CHARCO-10_Fo	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	2.5	31 Foqueo_F
CHARCO-10_Ge	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	57 Germinaci
CHARCO-10_Pri	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	6 Primera_F
CHARCO-10_Se	1	CHARCO-10	12/6/2014	4/26/2015	Silvio_Valle	12.3	7 Segunda_F

Cronologico **ProdAplicados** MonitoreoPLAGAS



Archivos desagregados y en otros formatos

Coordenadas mal registradas

37	5°08'27.5"	-75°54'31.3"	RISARALDA	APIA
38	5°08'42.3"	-75°55'02.2"	RISARALDA	APIA
39	5°78'41.0"	-75°05'02.8"	RISARALDA	APIA
40	5°67'16.8"	-75°04'17.0"	RISARALDA	APIA
41	5°08'17.8"	-75°54'18.4"	RISARALDA	APIA
42	5°09'41.8"	-75°55'26.4"	RISARALDA	APIA
43	5°09'41.6"	-75°55'26.1"	RISARALDA	APIA
44	5°09'35.3"	-75°55'10.0"	RISARALDA	APIA

<https://bigdata.cgiar.org/data-cleaning-the-unsexy-but-essential-aspect-of-data-science/>

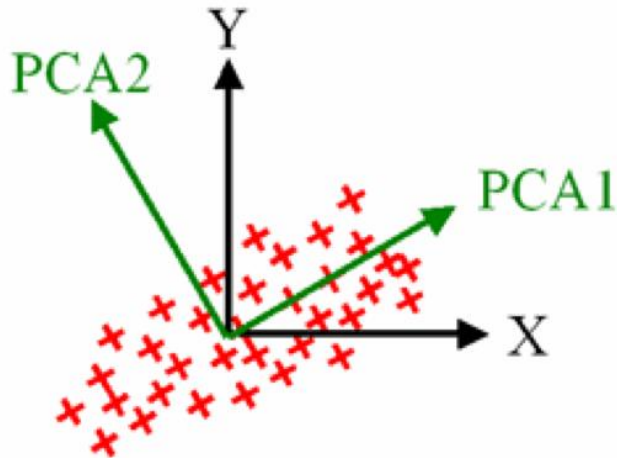
Date	value
19800101	NA
19800102	NA
19800103	NA
19800104	NA
19800105	NA
19800106	NA
19800107	NA
19800108	NA
19800109	NA
19800110	NA
19800111	35.2
19800112	NA
19800113	NA
19800114	36.2
19800115	35.2
19800116	NA

Valores perdidos

Métodos de análisis

Métodos tradicionales

- Regresión lineal múltiple (OLS)
- Análisis factoriales (PCA, MCA, CATPCA)
- Modelo lineal generalizado (GLM)
- Modelos mixtos



Métodos basados en aprendizaje automático

- Redes neuronales artificiales
- Random Forest - Conditional Inference Forest
- CART
- Support Vector Machine
- Optimización computacional



Diseminación de información a usuarios finales

Directamente hacia los productores por medios presenciales.

- Asistencia técnica
- Escuelas de campo (ECAs)
- Mesas técnicas agroclimáticas

Directamente hacia los productores por medios digitales

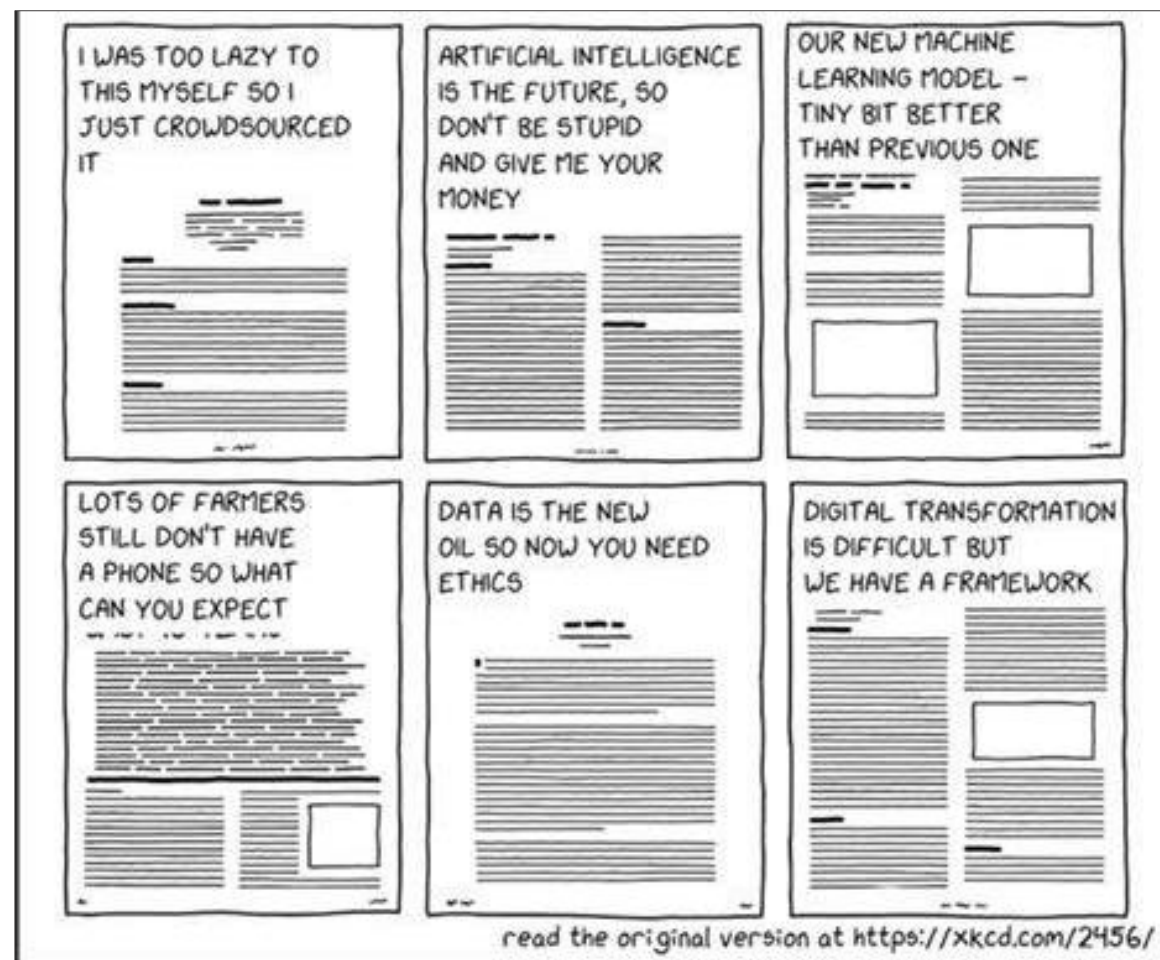
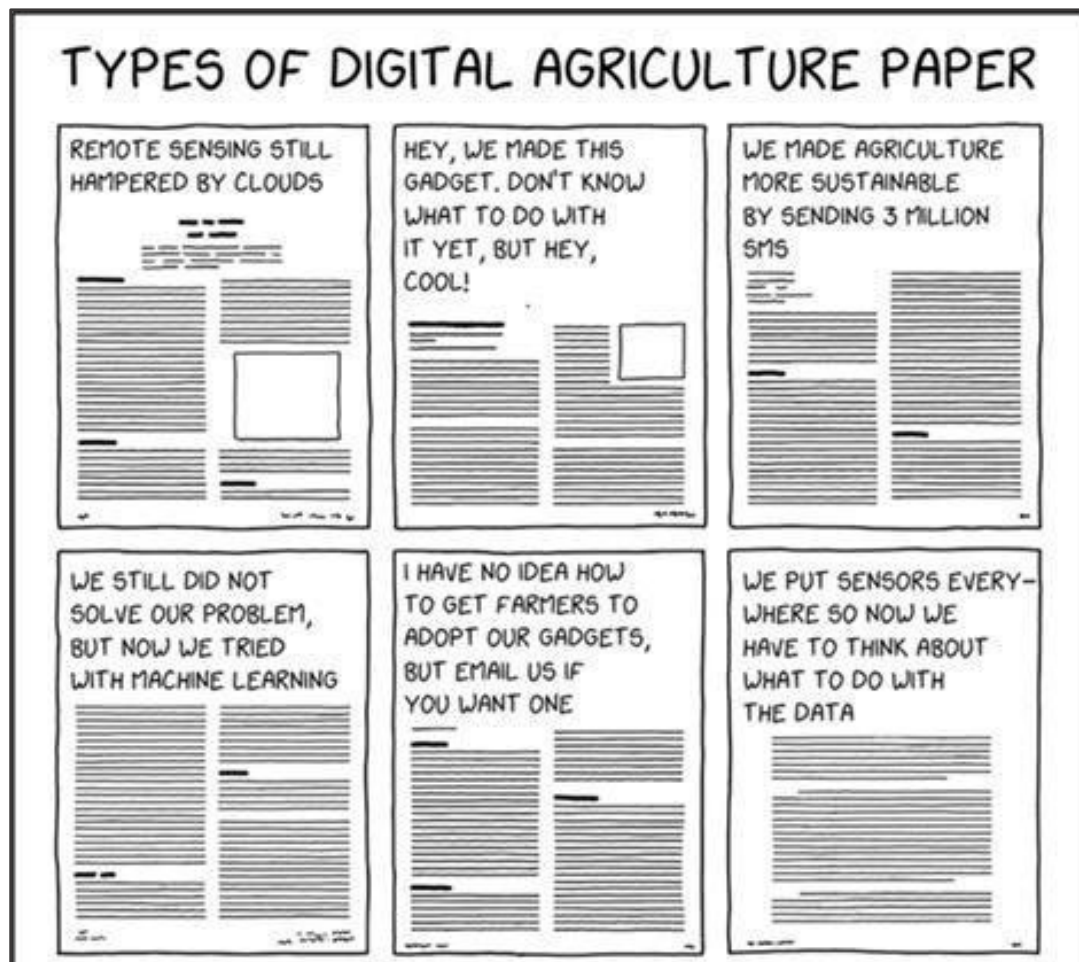
- Aplicaciones móviles.
- Chatbot. (Asistente de texto)
- IVR (Interactive voice response), llamadas
- Mensajes de texto

Otros usuarios (Gobierno, investigadores,...)

- Visualizaciones interactivas.
- Blogs.
- Notes de prensa.
- Publicaciones



Reflexión



Aspectos éticos

En el momento de capturar datos:

- Ley de habeas data - Derecho a conocer, actualizar, rectificar y suprimir los datos personales

En momento de publicar:

- Anonimizar información.
- Declarar una política de privacidad de los datos.



Herramientas y software

Procesamiento de datos



Datos espaciales

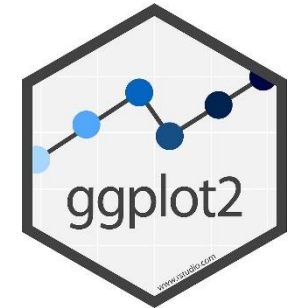


Quantum Gis

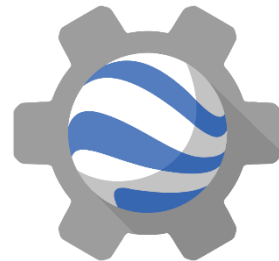
Computación en la nube



Visualización de datos



Análisis de datos



*Google earth
engine*



Recursos en línea.



<https://www.coursera.org/specializations/big-data>

<https://www.coursera.org/specializations/jhu-data-science>

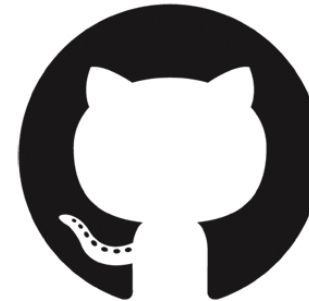


<https://www.edx.org/es/aprende/ciencia-de-datos>

<https://www.edx.org/es/micromasters/big-data>



<https://www.datacamp.com/courses/data-science-for-managers>



<https://github.com/hdorado>

<https://github.com/CIAT-DAPA>

Limitaciones

- El éxito de los resultados en el análisis de datos depende en gran parte de la **calidad de los datos**.
- Los resultados encontrados a partir de **los datos aplican solo para el universo de estudio**, lo cual los restringe a una población, un tiempo y ciertas condiciones.
- Algunos objetivos que podemos tener con el análisis de datos implica acceder **información que puede generar costo**.

Puntos claves

- La **importancia** de los **datos observacionales** en la investigación en agricultura.
- La existencia de **diversas fuentes** de **información** útiles en la agricultura.
- La ciencia de los datos involucra varios conceptos y disciplinas.
- El proceso de los datos para convertirse en información con la que se pueda tomar decisiones.
- La responsabilidad de la información que obtengamos, procesemos y produzcamos.
- Limitaciones.

Gracias!

Hugo Andrés Dorado.

Científico de datos

hugo.doradob@gmail.com

Conocimiento generado a partir de proyectos de:

Alianza

