M1 Panorámica de Data Science

Santander, 18 Septiembre 2017

OBJETIVOS Y METODOLOGÍA

Objetivos:

- Visión de global de Data Science (o Ciencia de Datos)
- Entender los problemas de Big Data
- Conocer el contexto de Open Science (o Ciencia Abierta)

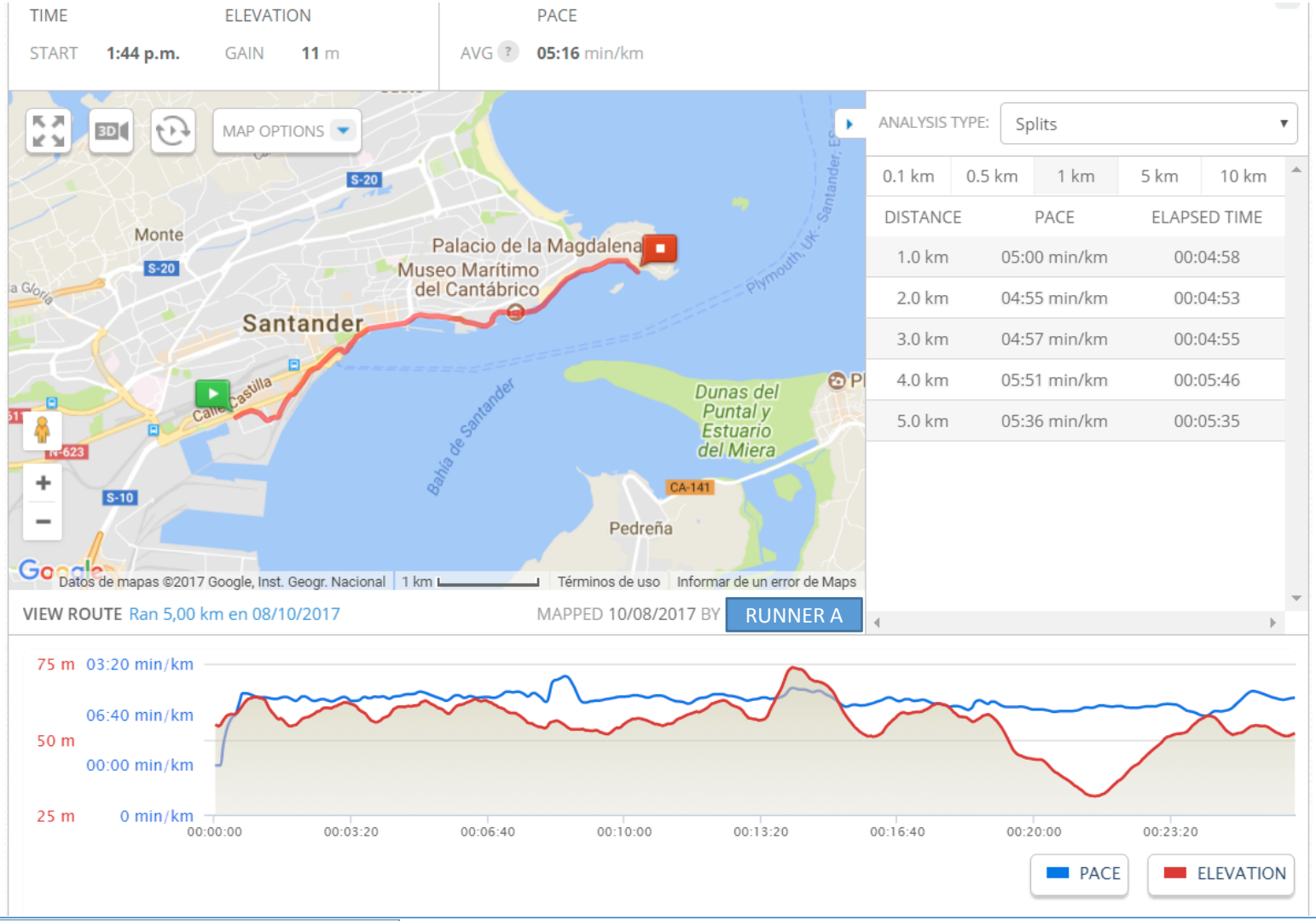
Metodología:

- Se comenzará por una exposición de los conceptos básicos, incluyendo ejemplos sencillos pero relevantes, que serán analizados individualmente y discutidos en común.
- Se revisarán los diferentes componentes de una solución, y los actores que participan en el desarrollo de la misma.
- Los estudiantes, organizados en grupos, realizarán un análisis detallado de un caso de estudio en una de las áreas comentadas (Salud, Medio Ambiente, Urbanismo, Economía, Astrofísica, Ciencias Sociales, Humanidades, Gestión Pública)

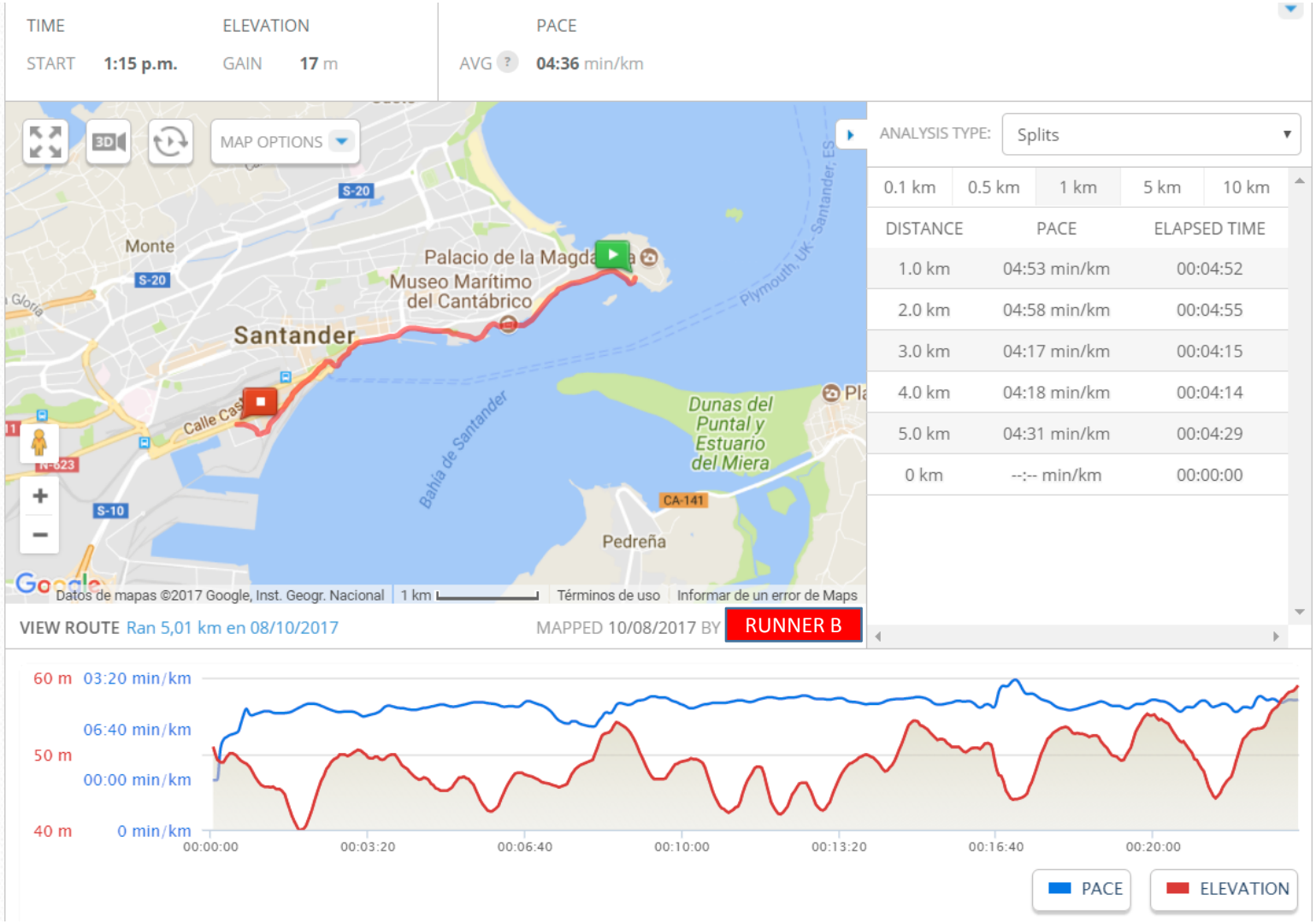
RESULTADOS DEL APRENDIZAJE

- Identificar problemas de Big Data en el contexto Open Science que requieren técnicas de Ciencia de Datos para su resolución
- Distinguir entre los diferentes componentes de dificultad de un problema de Big Data
- Saber elegir fuentes de información en el contexto de Open Science
- Identificar los elementos requeridos para abordar una solución completa, incluyendo la arquitectura y los diferentes actores implicados
- Conocer las nuevas técnicas en Data Science y las diferentes iniciativas en marcha e instituciones y empresas implicadas.
- Identificar el perfil profesional requerido en el contexto de Data Science

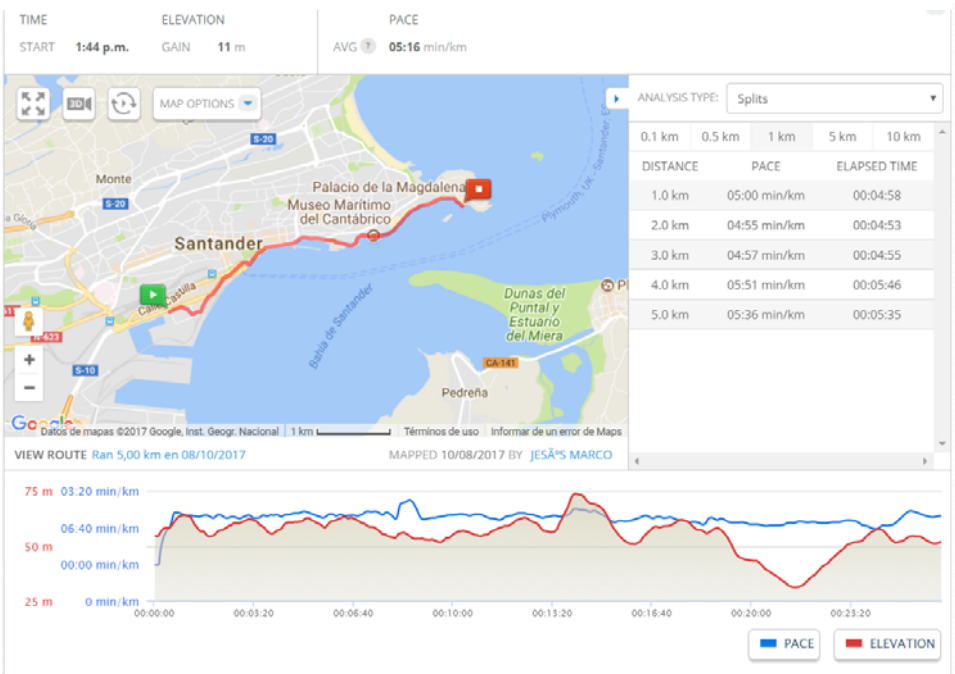
¿ENTENDER LOS DATOS ?

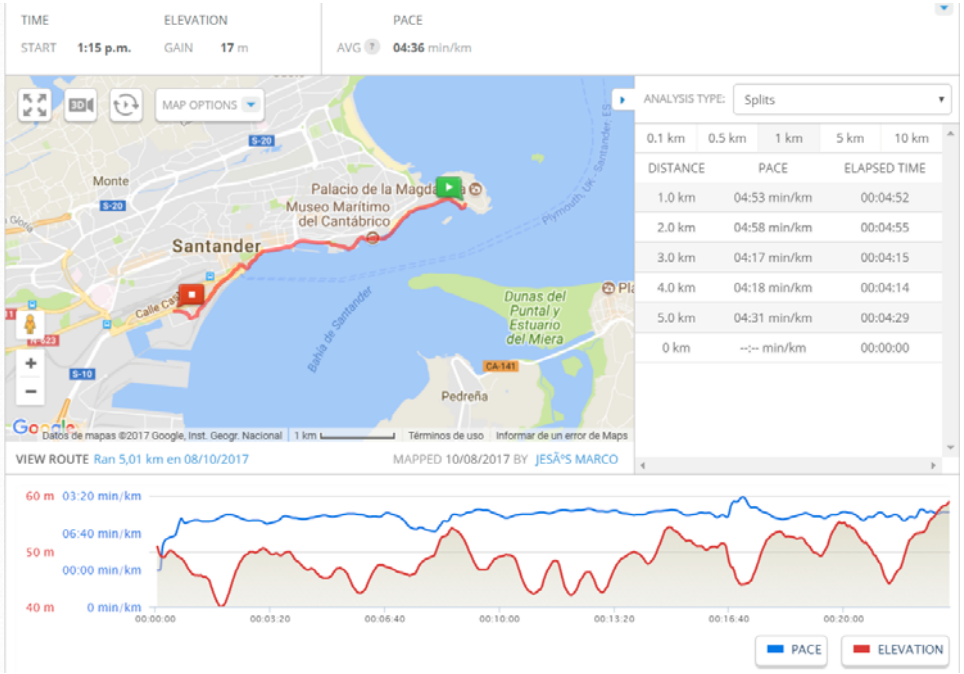


¿ENTENDER LOS DATOS?



DATOS Y… CONTEXTO: METADATOS !





-Cuestión 1: ¿cuál de los dos corredores, A o B, elegirías para tu equipo?
-Cuestión 2: ¿cómo influye la elevación del terreno?
-Cuestiones: ¿qué entendemos en base solamente a los datos?
¿Tenemos un "relato" propio o necesitamos un contexto?
Estos datos ¿deben estar en abierto?

COMPETENCIAS ESPECIFICAS

- DSDA-DA - Utilizar técnicas estadísticas apropiadas y análisis predictivo en los datos disponibles para descubrir nuevos patrones o relaciones.
- DS-ENG - Utilizar los principios de ingeniería para investigar, diseñar, desarrollar e implementar nuevos instrumentos y aplicaciones para la recogida de datos, su análisis y gestión.
- DS-RM - Crear nuevas visiones y capacidades mediante el uso del método científico (hipótesis, prueba, evaluación) que permitan descubrir nuevos enfoques para crear nuevos conocimientos, y lograr así los objetivos deseados en investigación u organización.
- DSRM01 - Crear nuevas visiones y capacidades mediante el uso del método científico (hipótesis, prueba y evaluación).
- DSRM02 – Realizar un estudio sistemático dirigido hacia un conocimiento más completo o la comprensión de los hechos observables, y descubrir nuevos enfoques para lograr los objetivos en investigación o de organización.

TEMARIO

- Introducción al contexto de Open Science (Ciencia en Abierto).
- Relevancia de los problemas de Big Data en Open Science
- El enfoque desde Ciencia de Datos
- Ejemplos relevantes de la aplicación de Ciencia de Datos en Open Science
- Arquitectura de las soluciones.
- Descripción de las e-Infraestructuras y el correspondiente Middleware
- Panorámica de casos de uso en las áreas de Salud, Medio Ambiente, Urbanismo, Economía, Astrofísica, Ciencias Sociales, Humanidades, Gestión Pública.
- Problemas y desafíos.
- Perfiles profesionales en el ecosistema de la Ciencia de datos.
- Nuevas tendencias y desarrollos.
- Proyectos e iniciativas, instituciones y empresas implicadas.

INTRODUCCION A OPEN SCIENCE

Objetivo*: lograr que la Ciencia sea mas *accesible*, eficiente y *transparente

***Impacto*: aprovechar mejor para la Sociedad los resultados que nos proporcionan los proyectos en Ciencia**

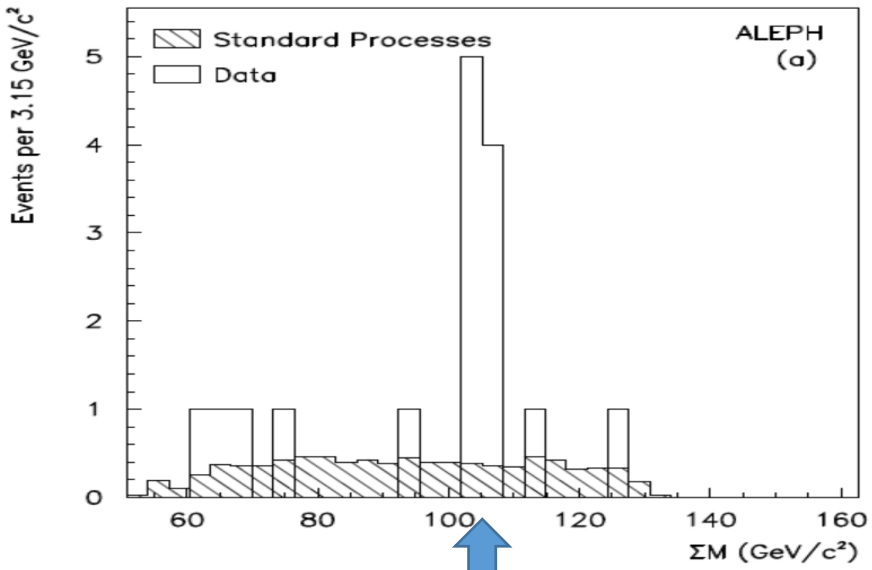
***Bonus*: hacer posibles nuevos desarrollos en Data Science**

Una experiencia personal

- ***HEP: we were (are?) so excellent and open…***

Story of a plot…

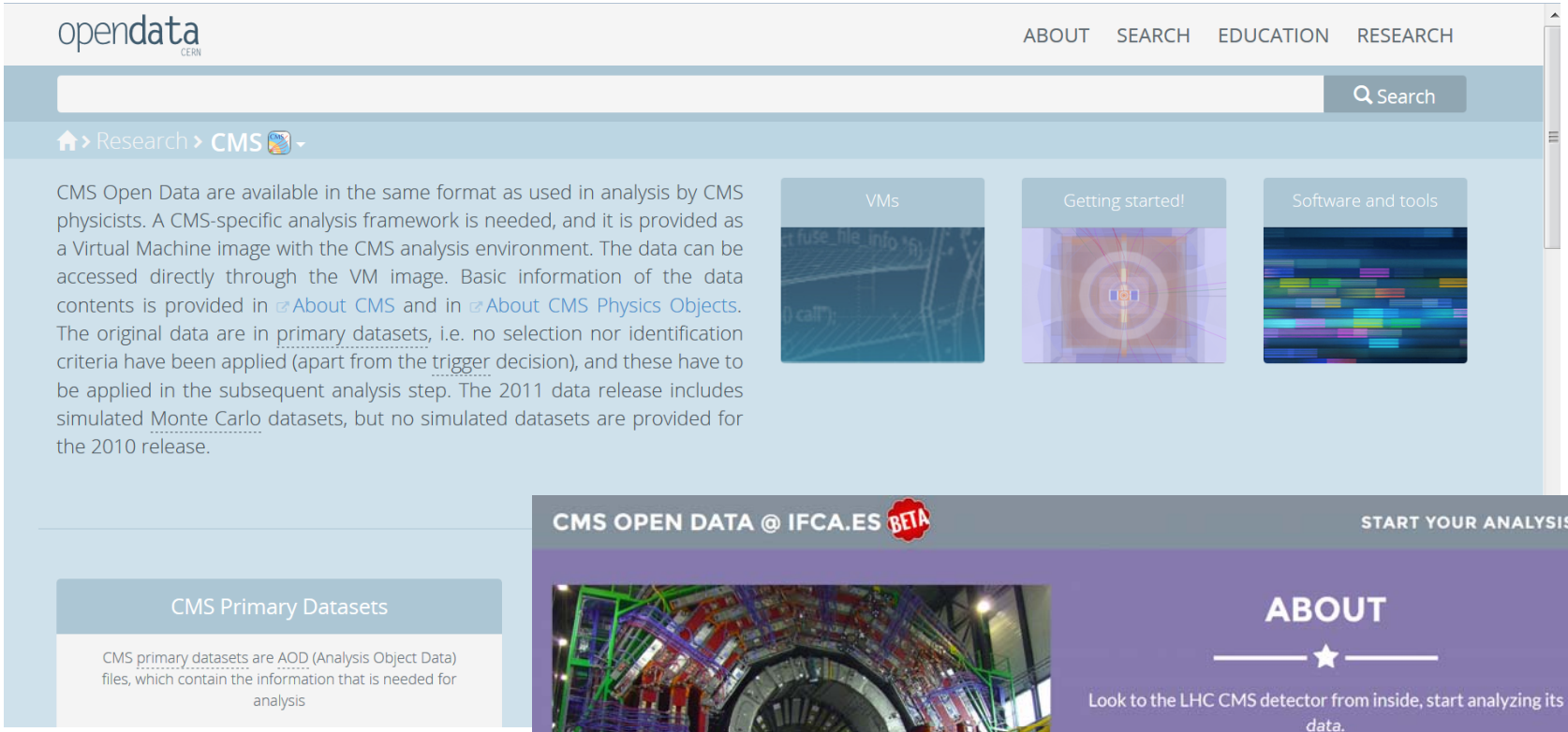
20 years ago!



Masa de una posible nueva partícula

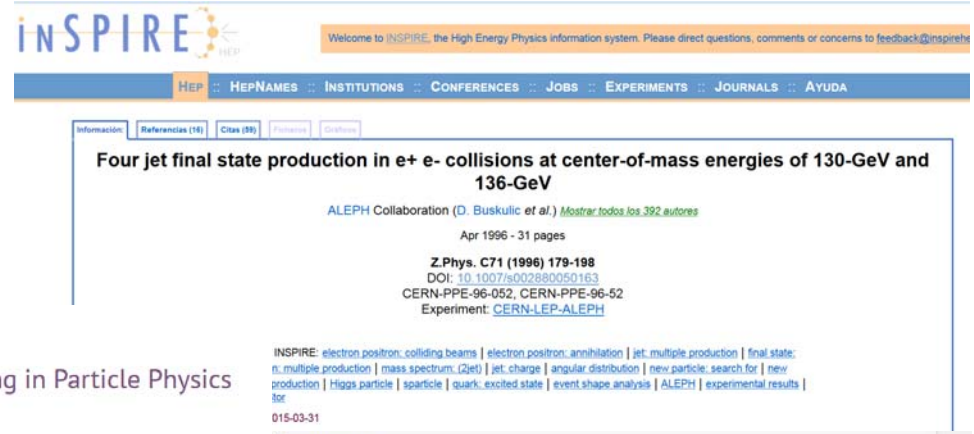
En Física de Partículas o en Astrofísica, tenemos Ciencia en Abierto?

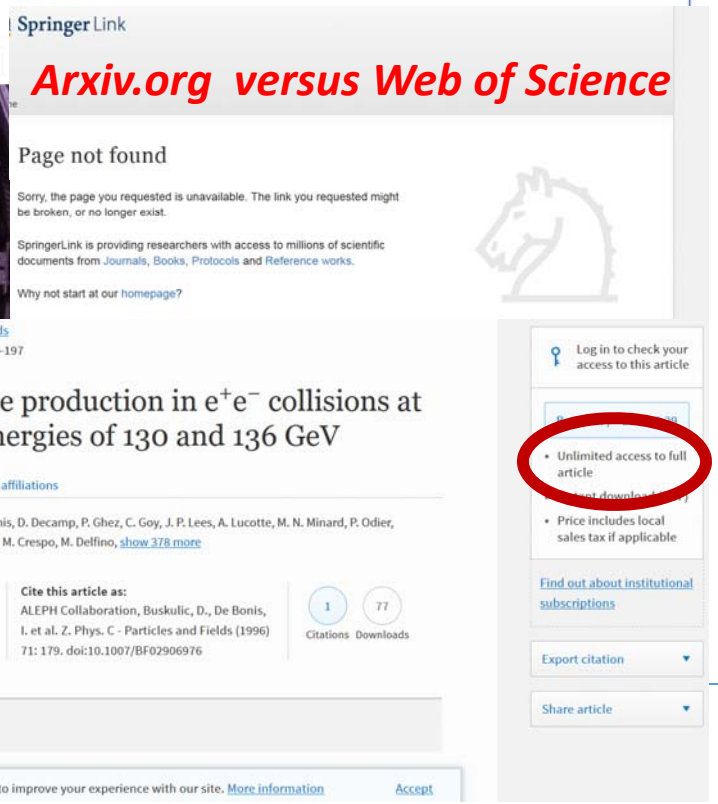
CMS OPEN DATA (2015)



Open Science en Física de Partículas

- ***HEP and Open Access***
  - ***CERN repositories***
  - ***SCOAP***





Arxiv.org versus Web of Science

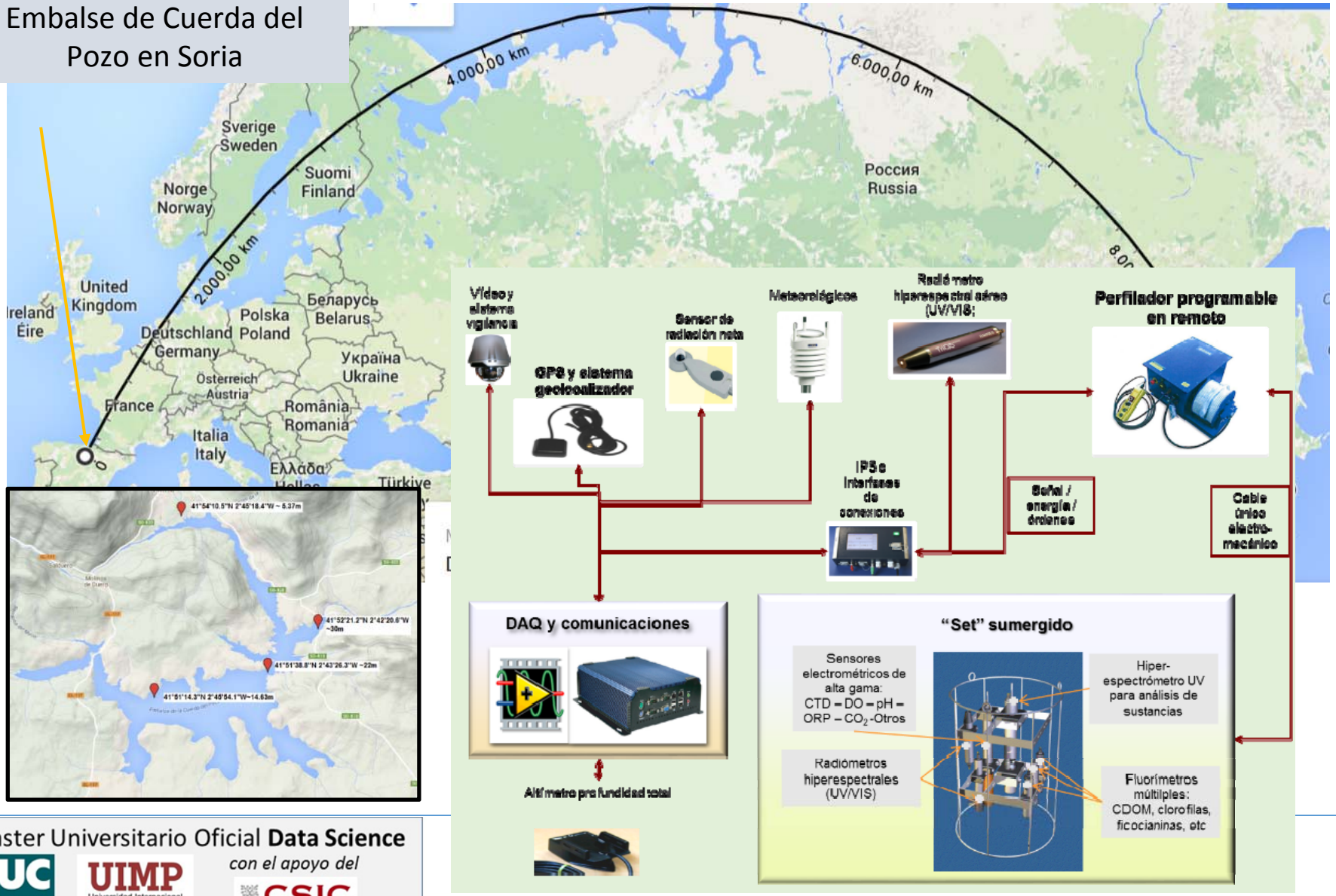
20 YEARS LATER…

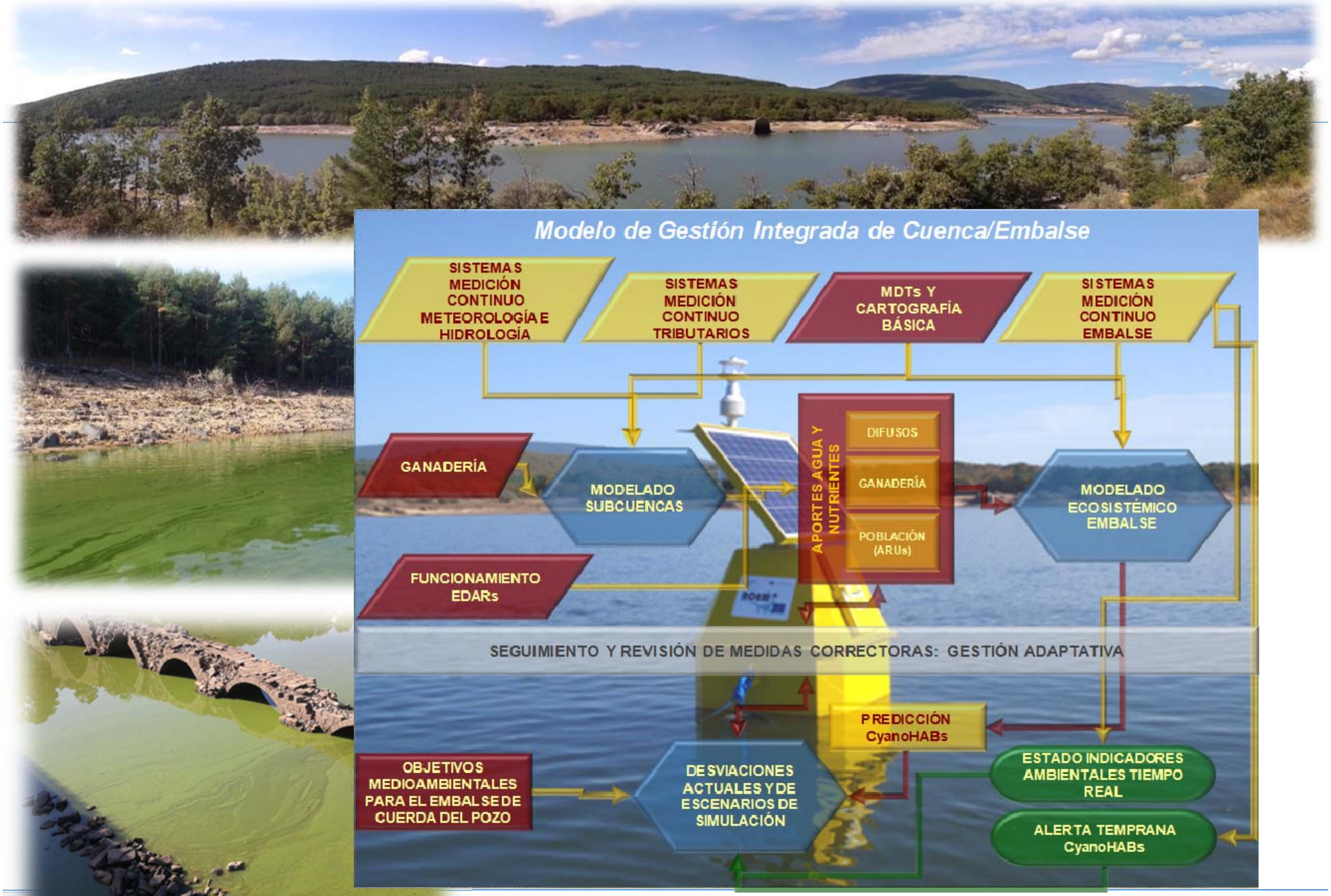
Ejercicio 1

- Analiza cómo problema de Open Science el descubrimiento del bosón de Higgs (2012)
  - Publicado en abierto
  - ¿Datos en abierto?
  - ¿Transparencia? ¿Reproducibilidad?
- Revisa el estado cómo problema de Open Science de la detección de ondas gravitacionales (Nobel 2017)
  - Publicado en abierto
  - ¿Datos en abierto?
  - ¿Transparencia?
- Deriva tus propias conclusiones iniciales

Open Science en un proyecto con una empresa: modelado de un embalse para entender eutrofización

Embalse de Cuerda del Pozo en Soria





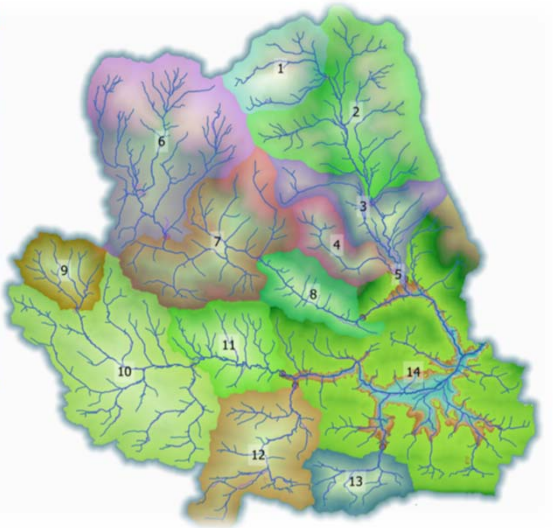
Ma

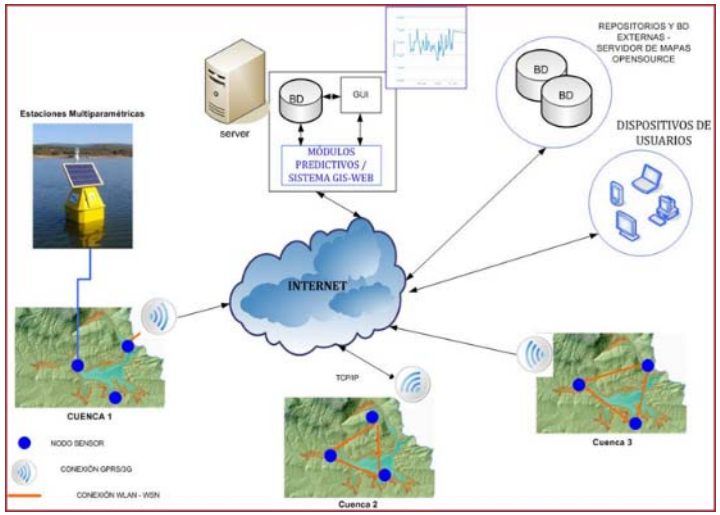
UC UNIVERSIDAD DE CANTABRIA · UIMP Universidad Internacional Menéndez Pelayo · con el apoyo del CSIC

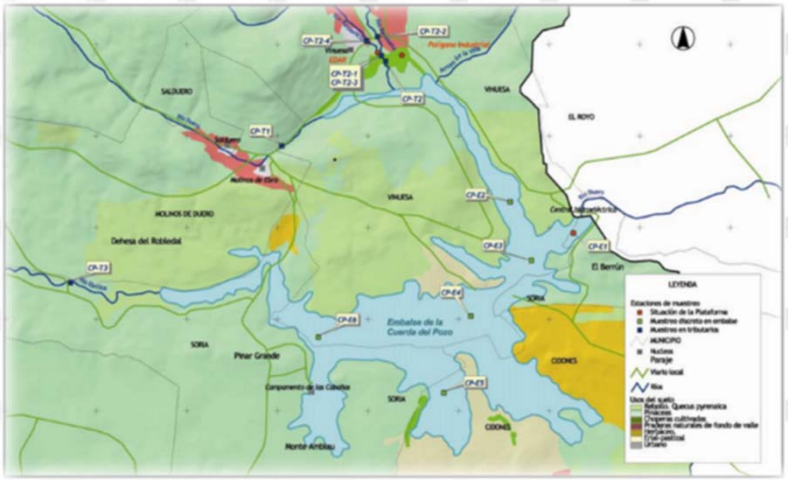
From A.Monteoliva PhD Thesis, 2016

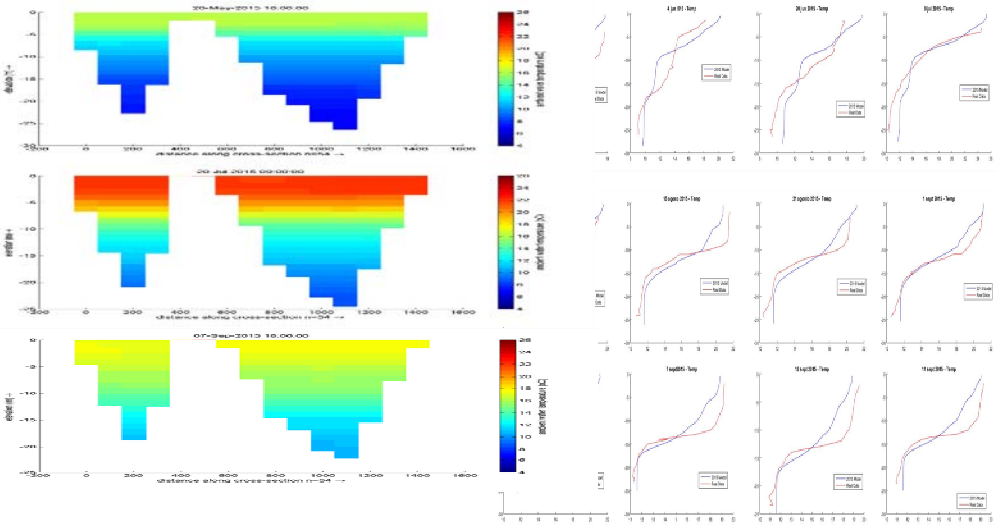
Modelo Hidrológico

ID	Nombre	AREA (Ha)
1	Revinuesa	2.142
2	Revinuesa	6.003
3	Revinuesa	2.530
4	Remonicio	1.397
5	Remonicio	13
6	Duero	6.720
7	Duero	4.878
8	Duero	1.810
9	Ebrillos	1.569
10	Ebrillos	7.249
11	Ebrillos	2.495
12	Dehesa	3.812
13	Bajero	1.615
14	CED	12.390

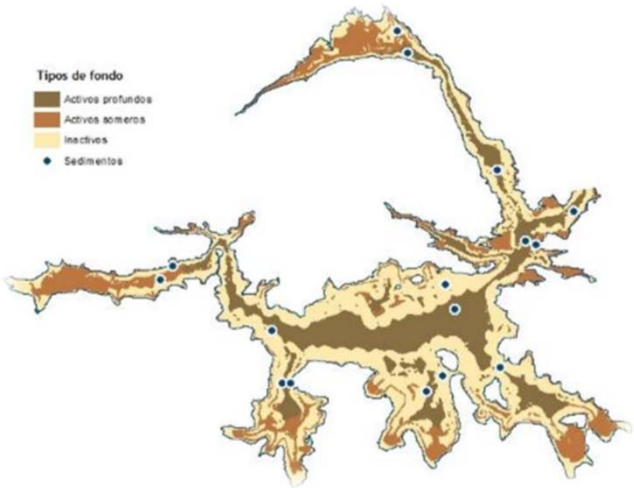


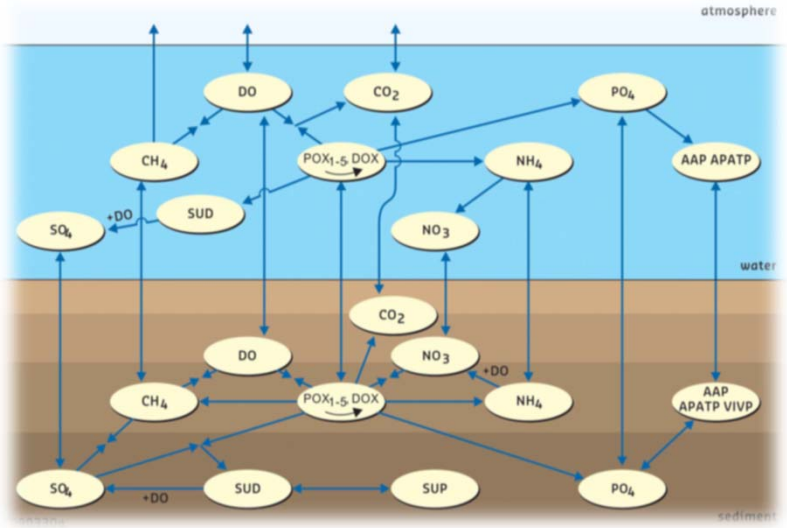




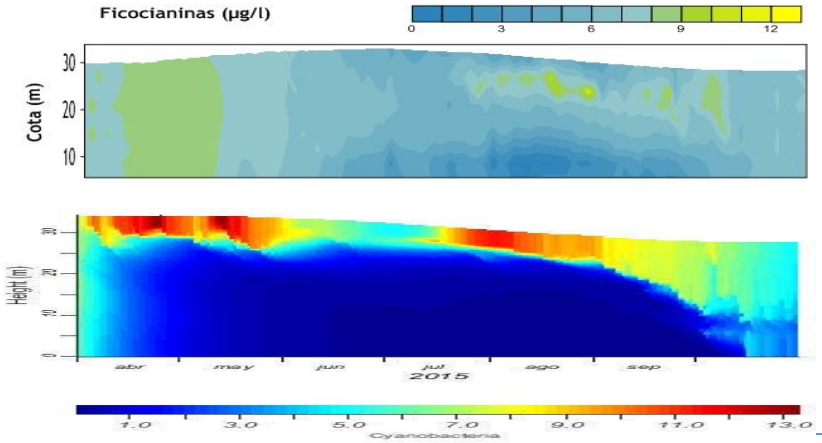


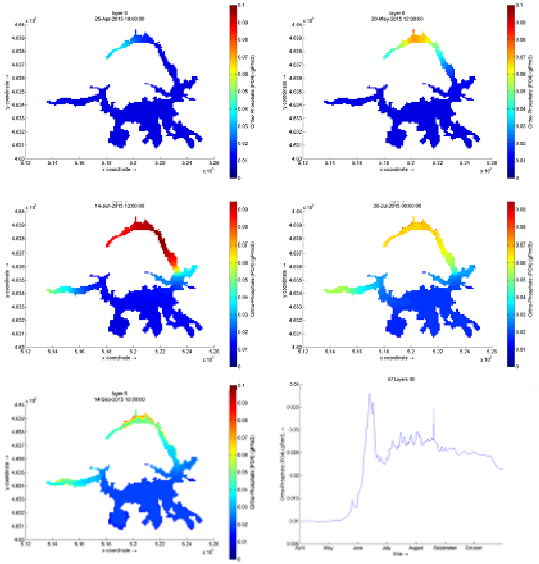
Modelo BioGeoQuímico:





Queremos predecir el posible crecimiento de algas tóxicas en el embalse



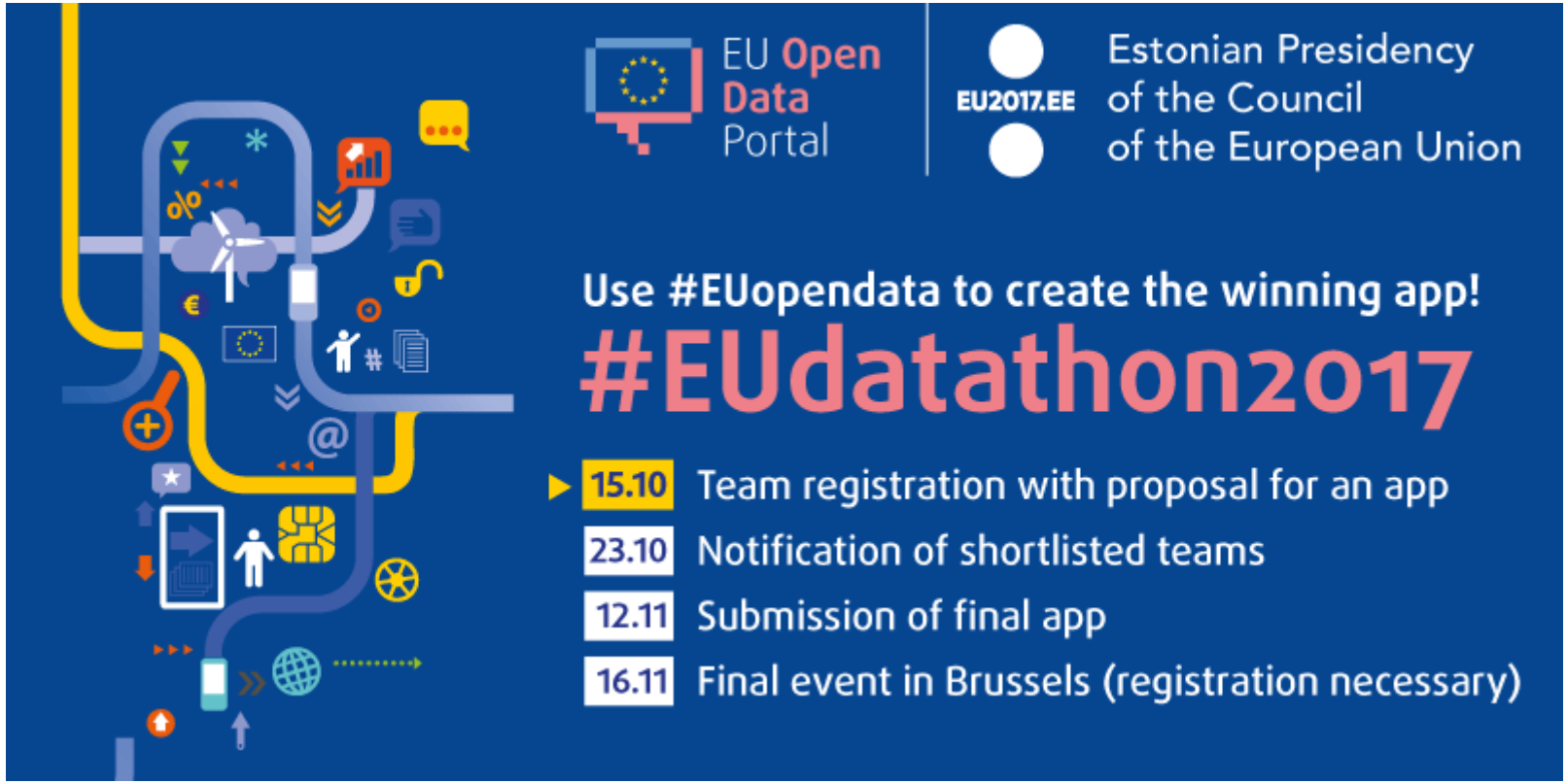


Ejercicio 2

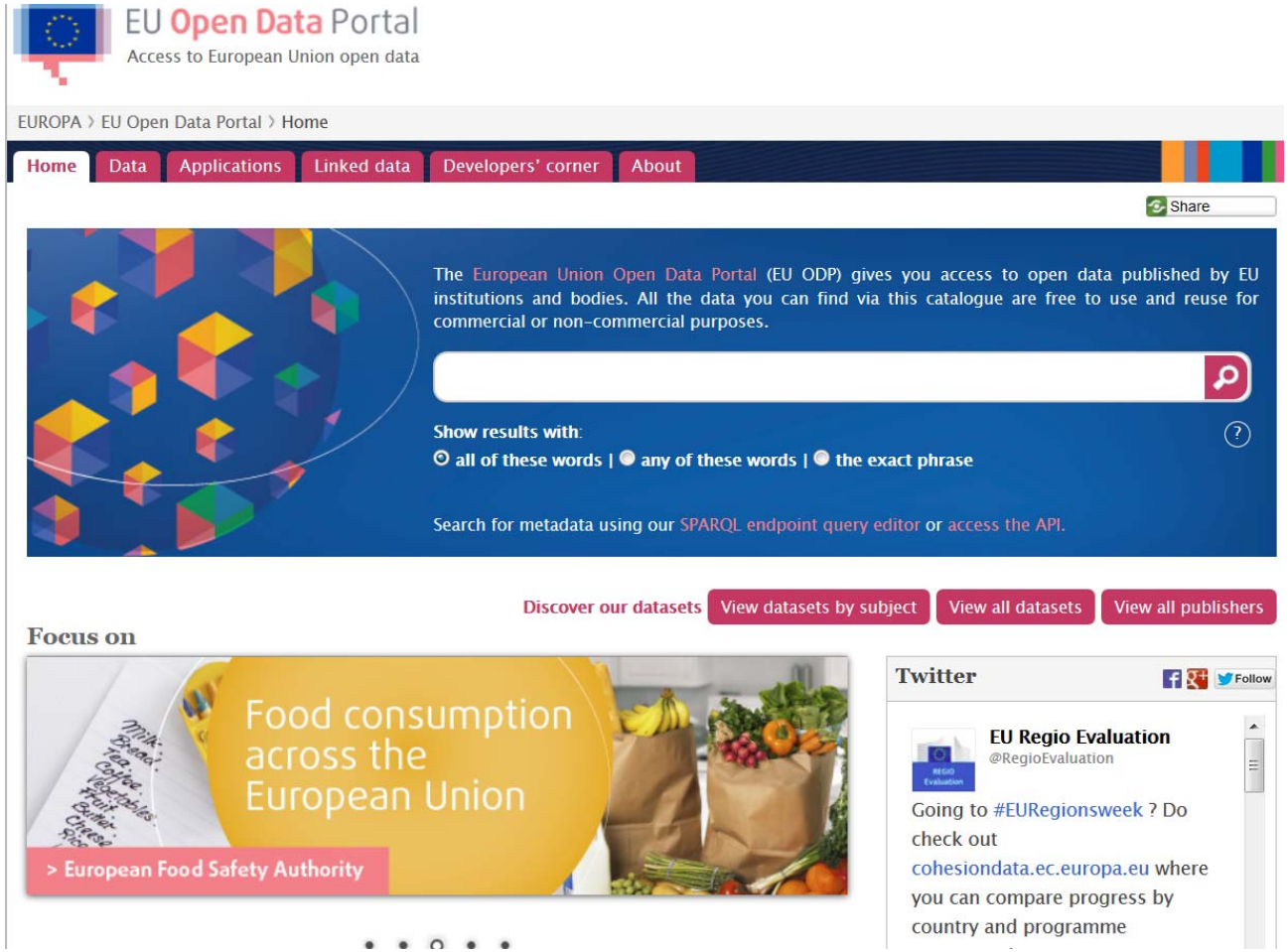
- Identifica el interés y los problemas de un enfoque “Open Science” para el modelado del embalse de Cuerda del Pozo
  - Datos externos para el modelo
  - Publicación de datos propios
  - Modelo en código abierto (open source)
  - Publicaciones del área en abierto
  - Validaciones en la literatura
  - Modelo de explotación de los resultados

El reto de la semana

- https://publications.europa.eu/en/web/datathon2017/home



Portal europeo de datos en abierto



https://data.europa.eu/euodp/en/home

Reto de la semana

- Identifica el interés de datos en abierto de instituciones públicas europeas para encontrar nuevas oportunidades en estas tres prioridades:

- Trabajo, crecimiento e inversión

- Mercado único digital

- Cambios democráticos

Tarea a completar

(deadline miércoles 20h)

Name of the team
Contact person
E-mail
Organisation (university, company, etc.)
Background (education, work, etc.)
Language of communication

Which EU dataset do you want to use (URL)?

Short description of the idea: **which question/challenge will you address?**

Short description of the idea: **how will you present the data?**

Short description of the idea: **how would you further develop this idea?**

¿¿¿ EXAMINAR DATASETS Y ENCONTRAR IDEAS
o
PENSAR EN IDEAS Y BUSCAR DATASETS ???