Portales y servicios de acceso a datos en abierto

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Másteres Universitarios

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.

DATOS GENERALES

Breve descripción

Esta materia tiene como objetivo proporcionar al estudiante un contacto directo con las tecnologías utilizadas para la implementación de repositorios de datos y su posterior explotación.

Se presentan los estándares más relevantes y las iniciativas más visibles en diferentes áreas de conocimiento dentro del contexto Open Science.

Título asignatura

Portales y servicios de acceso a datos en abierto

Código asignatura

102274

Curso académico

2020-21

Planes donde se imparte

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE

Créditos ECTS

4

Carácter de la asignatura

OPTATIVA

Duración

Cuatrimestral

Idioma

Castellano e Inglés

CONTENIDOS

Contenidos

1. Portales de acceso a datos en abierto.
2. Registro de datos y metadatos, estándares y herramientas.
3. Descarga de datos, acceso mediante web services y APIs.
4. Ejemplos relevantes de portales globales de datos en abierto (GEOSS).
5. Infraestructuras de procesado de datos en abierto.

RESULTADOS DE APRENDIZAJE Y DE FORMACIÓN

Generales

CG1 - Integrarse eficazmente en un grupo de trabajo y trabajar en equipo, compartir la información disponible e integrar su actividad en la actividad del grupo colaborando de forma activa en la consecución de objetivos comunes

CG6 - Buscar, obtener, procesar, comunicar información y transformarla en conocimiento

CG7 - Conocer las herramientas metodológicas necesarias para desarrollar proyectos avanzados

Transversales

CT1 - Analizar y combinar información utilizando diferentes fuentes

CT2 - Conocer la problemática éticas y legal relacionada con el análisis de datos y entender su importancia para una sociedad basada en los valores de la libertad, la justicia, la igualdad y el pluralismo

CT3 - Dominio de la gestión del tiempo

CT5 - Capacidad de trabajo autónomo y toma de decisiones

CT6 - Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, saber escuchar

Específicas

DSDM02 - Desarrollar e implementar modelos de datos, incluidos los metadatos

DSDM03 - Recoger e integrar diferentes fuentes de datos y su ingestión para su posterior análisis

DSDM05 - Asegurar la calidad de los datos, su accesibilidad, y su forma de publicación (curación)

DSENG01 - Aplicar los principios de ingeniería a la investigación, diseño y desarrollo de un prototipo de aplicaciones de análisis de datos, o al desarrollo de estructuras, instrumentos, máquinas, experimentos, procesos, sistemas requeridos para ello

DSENG02 - Desarrollar y aplicar soluciones computacionales para problemas en un cierto dominio de aplicación, usando una amplia gama de plataformas de análisis de datos

DSBPM02 - Utilizar los datos disponibles para mejorar los servicios existentes o desarrollar nuevos servicios

PLAN DE APRENDIZAJE

Actividades formativas

AF1 - Participación y asistencia a lecciones magistrales y seminarios

AF2 - Realización de prácticas de computación y análisis de datos

AF3 - Desarrollo de proyectos guiados

AF4 - Participación en casos prácticos en empresas o centros de investigación

AF6 - Tutorías (presenciales o por medio de recursos telemáticos)

AF7 - Elaboración de informes de laboratorio y trabajos

AF8 - Estudio individual de contenidos de la asignatura

AF9 - Trabajo en grupo

A10 - Pruebas de evaluación

Aprox. 100 horas:

- 30 % en clase
- 45% individual
- 25% en grupo

Metodologías docentes

En la asignatura 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.

Resultados de aprendizaje

- Identificar la forma de acceder a diferentes repositorios de datos en abierto para abordar un problema de Data Science.

- Implementar ejemplos sencillos pero útiles de acceso a portales de datos relevantes directamente o mediante APIs en un contexto de web services.
- Operar un repositorio sencillo con datos históricos. Implementar una versión ya disponible y proponer mejoras en el diseño del mismo.
- Ilustrar cómo el software existente puede usarse para mejorar la calidad de los datos, y en particular su accesibilidad.

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

SE1 - Examen (escrito, oral y/o práctico en el aula de computación) (ponderación mínima 0% y máxima 40%)

SE2 - Valoración de informes y trabajos escritos (ponderación mínima 0% y máxima 60%)

SE3 - Valoración de exposiciones orales de trabajos (ponderación mínima 0% y máxima 60%)

SE4 - Seguimiento de actividades presenciales (ponderación mínima 0% y máxima 40%)

PROFESORADO

Profesor responsable

Bernal Martínez, Isabel

Responsable General de Digital
Unidad de Recursos de Información Científica para la Investigación
Consejo Superior de Investigaciones Científicas (CSIC)

Profesorado

Profesor Responsable de la asignatura

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Science as an Open Enterprise (2012): https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/

Open Archives Initiative: http://www.openarchives.org/

Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers: http://roytennant.com/bitter_harvest.html

DataCite Content Negotiation: https://support.datacite.org/docs/datacite-content-resolver

Metadata harvesting through schema.org: https://figshare.com/articles/Metadata_harvesting_through_schema_org/9120509/2

DCAT Application Profile for data portals in Europe: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/about

Data aggregators: a solution to open data issues: https://blog.okfn.org/2017/12/28/data-aggregators-a-solution-to-open-data-issues/

OpenAire Aggregation and content provision workflows: https://www.openaire.eu/aggregation-and-content-provision-workflows

Open Access Tracking Project (OATP): http://tagteam.harvard.edu/hubs/oatp/tag/oa.repositories.data

Open Data Handbook http://opendatahandbook.org/

Recommendations for Open Data Portals: from setup to sustainability (2020): https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf

Recommendations For Services In A FAIR Data Ecosystem (2020): https://www.rd-alliance.org/recommendations-services-fair-data-ecosystem-0

Open Standards for Data: https://standards.theodi.org/find-existing-standards/, https://standards.theodi.org/introduction/types-of-open-standards-for-data/

PID Services Registry: https://pidservices.org/

The Right to Read is the Right to Mine:

- https://sparcopen.org/news/2015/the-right-to-read-is-the-right-to-mine/#:~:text=%E2%80%9CThe%20Right%20to%20Read%20is%20the%20Right%20to,to%20perform%20research%20using%20humanity%E2%80%99s%20accumulated%20scientific%20knowledge.%E

2%80%9D

- https://blogs.lse.ac.uk/impactofsocialsciences/2014/06/04/the-right-to-read-is-the-right-to-mine-tdm/
- https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/

Dramatic Growth of Open Science Portal: https://poeticeconomics.blogspot.com/

Data Portals: http://dataportals.org/

EOSC Portal: https://www.eosc-portal.eu/, https://marketplace.eosc-portal.eu/

Datos.gob.es: https://datos.gob.es/en

Developing an Open Data Portal for the ESA Climate Change Initiative (2020): https://datascience.codata.org/articles/10.5334/dsj-2020-016/

Navigating the unfolding open data landscape in ecology and evolution (2018): https://core.ac.uk/download/pdf/189497983.pdf

A survey of Web technology for metadata aggregation in cultural heritage (2018): https://content.iospress.com/articles/information-services-and-use/isu859

Open Data Best Practices in Europe: https://www.europeandataportal.eu/en/highlights/open-data-best-practices-europe

Data management and use: Governance in the 21st century - a British Academy and Royal Society Project: https://royalsociety.org/topics-policy/projects/data-governance/

Privacy Enhancing Technologies: https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/