



UNIVERSIDAD NACIONAL DE SAN LUIS
ARGENTINA

HACIA LA EVALUACIÓN DE LA CALIDAD DE DATOS ABIERTOS

Ana Funes¹, Aristides Dasso¹, María Alejandra Barrera²

¹ Departamento de Informática, Facultad de Ciencias Físico-Matemáticas y Naturales,
Universidad Nacional de San Luis
Ejército de los Andes 950, D5700HHW San Luis, Argentina
afunes@email.unsl.edu.ar , arisdas@unsl.edu.ar

² Departamento de Informática, Facultad de Tecnología y Ciencias Aplicadas,
Universidad Nacional de Catamarca,
Maximio Victoria 55, Catamarca, Argentina
mbarrera@tecno.unca.edu.ar

CONTEXTO

Investigación conjunta entre investigadores de los Proyectos PROICO 03-2020 “Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube”, de la Universidad Nacional de San Luis e investigadores del Proyecto: “Ingeniería de Software en la era de las Industrias 4.0.” de la Universidad Nacional de Catamarca.

LÍNEAS DE I + D

Creación de un modelo de atributos de calidad de (DA) y la definición, sobre la base de los atributos identificados, de métricas que sirvan para obtener un indicador de la calidad de los mismos.

RESUMEN

Según la Open Knowledge Foundation, ‘datos abiertos’ son datos que están disponibles de forma pública a cualquier persona que desee utilizar y redistribuir libremente, sujetos al requisito de atribuir y compartir por igual. Están disponibles en formatos digitales y publicados bajo diversas licencias. Dado los beneficios de su adopción, hay múltiples iniciativas que buscan promover su uso, p.e. Open Government Partnership, el Portal de Datos Abiertos de la UE, el Portal de Datos Abiertos de los EEUU, etc. En Argentina, Datos Argentina tiene como objetivo mejorar la transparencia y la participación ciudadana mediante la publicación de datos abiertos. Frente a este escenario, consideramos importante contar con modelos y herramientas que permitan evaluar la calidad de los conjuntos de datos publicados o a publicar.

RESULTADOS Y OBJETIVOS

El principal objetivo es la construcción de un modelo basado en el método de evaluación multicriterio Logic Scoring Preferences (LSP), con el objeto de medir y evaluar de manera integral y parcial el grado de cumplimiento de determinados atributos y sub-atributos de calidad de Datos Abiertos (DA).

Se planea (a) Desarrollar un modelo jerárquico de atributos de calidad de datos abiertos, que unifique e incluya las dimensiones, subdimensiones y atributos necesarios para definir una métrica de calidad que considere estos aspectos. Varios de estos atributos han sido identificados también en otras propuestas e iniciativas, pudiendo observar que muchas veces se consideran atributos de calidad comunes aunque bajo diferentes nombres. (b) A partir del modelo jerárquico de calidad construido, definir un conjunto de métricas que permitan medir los atributos hojas de dicha jerarquía. (c) Definir una función multicriterio que agregue las características antes identificadas y que permita la evaluación tanto integral como de subcaracterísticas del modelo jerárquico. Este modelo cuantitativo de evaluación, permitirá obtener los valores de de las características evaluadas a diferentes grados de abstracción así como un valor de un indicador global del repositorio de datos evaluado. (d) Aplicar el modelo y las métricas asociadas para evaluar la calidad de los datos a diversos portales de datos abiertos.

FORMACIÓN DE RECURSOS HUMANOS

La investigación y construcción de modelos de evaluación de sistemas complejos ya ha producido varias tesis de posgrado en la Universidad Nacional de San Luis (UNSL). Uno de los autores de este trabajo ha realizado su tesis de Maestría en Ingeniería de Software de la UNSL sobre la evaluación de la reusabilidad de datos espaciales abiertos. Esta propuesta tiene como objetivo también continuar con la formación de RRHH tanto en el Dpto. de Informática de la UNSL como en el Dpto. de Informática de la Universidad Nacional de Catamarca.