

Tesis doctoral de María Inmaculada Aleixos Borrás

Título de la tesis: A causal model to explain data reuse in science: a study in health disciplines

Resumen en español:

Las inversiones en infraestructuras de datos, gestión de datos, repositorios de datos y políticas y recomendaciones de intercambio de Datos Abiertos (*Open Data*) se consideran cada vez más importantes para la producción del conocimiento científico. Una de las razones que justifica estas inversiones es que cuanto más Datos Abiertos haya, mayores serán las posibilidades de crear nuevo conocimiento que pueda hacer avanzar tanto la ciencia como el bienestar humano. Sin embargo, los esfuerzos y la inversión en Datos Abiertos y otras formas de compartirlos sólo tienen valor si se reutilizan realmente. Recientes trabajos académicos han puesto de manifiesto algunos de los retos y factores facilitadores relacionados con la reutilización de los datos, a fin de asesorar las políticas e inversiones actuales y futuras. Sin embargo, a pesar de esos esfuerzos, todavía desconocemos por qué y cómo algunos/as investigadores/as logran reutilizar los datos, a pesar de los retos a los que enfrentan, y por qué otros/as investigadores/as abandonan el proceso de reutilización de los datos. La presente tesis tiene por objeto llenar este vacío centrándose en una explicación causal del proceso de reutilización de los datos, que se entiende está inmersa en pautas de conducta más amplias que se relacionan con las motivaciones, los objetivos científicos y las estrategias de toma de decisiones de los/as investigadores/as.

Esta tesis consta de tres elementos principales. En primer lugar, propone un modelo heurístico del actor científico, el *modelo del horizonte individual delimitado* (BIH por el nombre inglés, *bounded individual horizon*). En él se entiende que, por una parte, el trabajo y la carrera de los/as investigadores/as se estructuran en función de su motivación para producir contribuciones científicas y de los sistemas de recompensa que dan prioridad a determinados tipos de contribuciones. Por otra parte, los esfuerzos de los/as investigadores/as para lograr su objetivo de crear nuevos hallazgos que acumulen reconocimiento y recompensas se producen en un marco de información y recursos limitados, condicionados por múltiples factores institucionales, sociales y de otra índole. En segundo lugar, esta tesis propone una explicación teórica causal mecanicista que permite comprender el proceso de reutilización de los datos y sus efectos (resultados). El *mecanismo de reutilización de datos* (*data-reuse mechanism*), como se denomina, nos permite comprender cómo la toma de decisiones científicas está caracterizada por una conducta que tiende a satisfacer esos objetivos en unas condiciones y procesos específicos de reutilización de datos. En tercer lugar, esta tesis incluye los resultados del estudio empírico de diez estudios de casos de reutilización de datos en ciencias de la salud. Estos casos se han analizado e interpretado utilizando el modelo teórico del *horizonte individual delimitado* y los enfoques del *mecanismo de reutilización de datos*.

Los resultados principales explican que existe una aparente asociación entre el alcance el alcance y tipo de esfuerzo requerido para reutilizar datos, las motivaciones contextualizadas de los/as

investigadores/as y marcos más amplios de fijación de objetivos y toma de decisiones. El acceso a los datos es una condición necesaria para su reutilización, pero no es suficiente para que ésta se produzca. Para comprender por qué algunos procesos de reutilización de datos persisten y tienen éxito, mientras que otros no, son elementos esenciales: las características de los datos disponibles, incluido el contexto de su producción; el grado de preparación y administración de esos datos; y su potencial valor en relación con las motivaciones de los investigadores para hacer nuevas afirmaciones científicas o generar conocimientos de base. Este estudio concluye que los esfuerzos e inversiones destinados a aprovechar los beneficios de la reutilización de los datos también deberían ampliarse para incluir la capacitación de los/as investigadores/as en materia de reutilización de datos. En particular, debe insistirse en la capacidad para reconocer eficientemente las oportunidades, sortear los problemas del proceso de reutilización y ser conscientes y reconocer las limitaciones de la utilización de datos secundarios. Sin estas inversiones, las promesas y expectativas vinculadas a las emergentes infraestructuras de datos, los repositorios de datos, las directrices de gestión de datos y las prácticas científicas abiertas tienen muchas menos probabilidades de alcanzar su pleno potencial.