

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática *

Mariano Competella¹, Alejandra Cechich¹, Agustina Buccella¹,
Ayelén Montenegro², Ángel Muñoz², and Andrea Rodríguez²

¹ GIISCO Research Group

Departamento de Ingeniería de Sistemas - Facultad de Informática

Universidad Nacional del Comahue, Neuquen, Argentina

mariano.competella@est.fi.uncoma.edu.ar,

alejandra.cechich,agustina.buccella}@fi.uncoma.edu.ar

² Instituto Nacional de Tecnología Agropecuaria (INTA)

Alto Valle de Río Negro y Neuquén

montenegro.ayelen@inta.gob.ar

Abstract. Considerando la cantidad y diversidad en los datos que hoy día se relevan para futuros análisis, su combinación y uso se torna un elemento complejo a modelar. Es por esto que el agregado de semántica, a través de modelos conceptuales, es una tendencia actual en las arquitecturas software para Sistemas Big Data. En ese sentido, en este artículo presentamos una caracterización de contexto mediante la identificación top-down de variedad en sistemas predictivos sobre fluctuaciones de cuerpos de aguas subterráneos. Esa caracterización favorecería la identificación de situaciones recurrentes, incluyendo la posibilidad de reusabilidad durante el análisis. La propuesta se ejemplifica mediante dos casos comparativos en zonas geográficas diferentes y distantes.

Keywords: Reusabilidad de Software · Variedad en Sistemas Big Data · Taxonomías

1. Introducción

Nuestra propuesta para modelar variedad en sistemas big data (SBDs) se ve influenciada por distintos tipos de diversidad [1]: de *fuentes*, detectando distintas estructuras en los datos, técnicas de adquisición, etc.; de *contenido*, enfocando en las variables que son relevantes a un determinado problema de dominio; y de *procesamiento*, detectando variaciones en técnicas de análisis. Sin embargo, la variedad de los datos ha sido también considerada desde el punto de vista de incorporar semántica al proceso de modelado, incluso relacionando propiedades de calidad como interoperabilidad, seguridad, reusabilidad, etc. Entonces, en trabajos previos, hemos redefinido y extendido la clasificación de variedad para incluir la *variedad de contexto* [9], definiéndola como aquella que se refiere a *variaciones de dominio que pueden restringir o afectar los resultados del análisis*. Los diferentes tipos de contexto dependen de cada dominio; por ejemplo, para sistemas hidrológicos, los cuerpos de agua pueden clasificarse en ríos, lagos, mares, etc., que pueden verse influenciados por

* Este trabajo esta parcialmente soportado por el proyecto UNCOMA 04/F019 “Variedad en Sistemas Big Data” 2022-2025.

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*

la climatología, geología, u otros factores de entorno. Por lo tanto, estructurar conocimiento de dominio es una tarea esencial para detectar la variedad contextual.

El proceso de identificación de variedad en SBDs puede iniciarse desde la evidencia que presenten los datos en datasets (enfoque bottom-up); desde el análisis conceptual del dominio (enfoque top-down); o con una mezcla de ambos casos. En [9] esos enfoques se discuten con cada uno de sus pasos, incluyendo la caracterización de dominio que deriva en la construcción de artefactos mediante los cuales se resguardan variables y relaciones relevantes para describir contextos similares. Esos artefactos son gestionados por una herramienta de soporte (CoVaMaT), que permite almacenarlos con una estructura que facilite su recuperación y reuso [10].

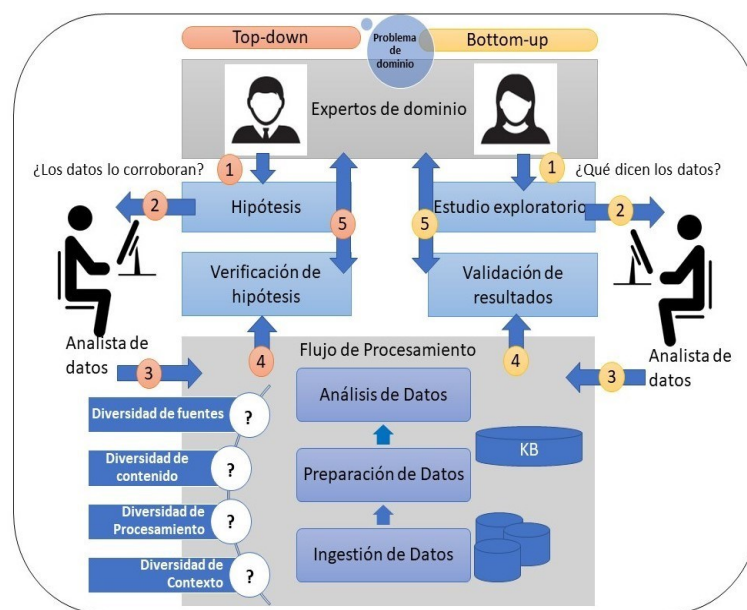


Fig.1. Enfoques T-VIP y B-VIP para identificación de variedad [9]

La Figura 1, muestra los dos enfoques paso a paso:

- Enfoque top-down (T-VIP): en este enfoque, a la izquierda de la figura, (1) dado un problema de dominio, el usuario experto elabora una o más hipótesis que deberán ser probadas mediante análisis de datos (¿los datos lo corroboran?); (2) luego, en base a esas hipótesis, analistas de datos proceden a realizar las tareas del flujo de procesamiento (3); finalmente, los resultados permiten verificar las hipótesis (4) posiblemente visualizando los datos de diferentes maneras, incluso permitiendo reformulaciones (5).
- Enfoque bottom-up (B-VIP): en este enfoque, a la derecha de la figura, (1) dado un problema de dominio, el usuario experto decide iniciar un estudio exploratorio para averiguar lo que los datos pueden revelar sobre ese problema (¿qué dicen los datos?); (2) luego, el estudio es llevado a cabo por analistas de datos aplicando las

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*

actividades del flujo de procesamiento (3); finalmente, los datos retornan (4) para ser validados por expertos, finalizando el proceso o reiniciando la búsqueda (5).

La Figura 1 resalta activos en una base de conocimiento (KB). La forma en que esos artefactos se identifican varía dependiendo del enfoque. Por ejemplo, en el caso top-down, implica entender e investigar las características del dominio, probablemente estructurando el conocimiento como taxonomías, antes de proceder a contextualizar y almacenar cada caso de estudio para su reuso futuro. En este artículo, presentamos una caracterización top-down realizada para el reuso de sistemas predictivos sobre un dominio concreto: las fluctuaciones de napas freáticas influenciadas por condiciones meteorológicas y de entorno. Las variables relevantes del dominio son organizadas en forma de una taxonomía, que luego es instanciada en dos casos de estudio a modo comparativo para destacar la influencia de ese entorno en posibilidades de reuso de sistemas big data.

El artículo se organiza de la siguiente manera. La sección siguiente describe la caracterización de dominio realizada en el enfoque top-down para el problema bajo estudio. Luego, se analizan dos casos de estudio de manera comparativa y a modo de prueba de conceptos. Conclusiones y trabajos futuros se abordan al final.

2. Caracterización de influencias en la predicción de fluctuaciones de la napa freática

La Figura 2, muestra el proceso T-VIP para el caso de estudio comparativo. Note que se accede a la base de soporte gestionada por CoVaMaT para ingresar, consultar o reusar casos de variedad a modo de hojas funcionales* [9]. Las siguientes secciones, describen el proceso paso a paso.

2.1 T-VIP paso a paso: Identificación de variables

(1) *El problema de dominio.* Como puede verse en la figura, el problema de dominio se refiere al análisis de factores que se relacionan con fluctuaciones de cuerpos de agua subterráneos o napas freáticas. El uso de big data en este campo nos permite saber dónde ocurrirán las mayores fluctuaciones para brindar, a quienes deban tomar decisiones, una idea de qué áreas requerirán más atención y dónde deberían asignarse recursos más adecuadamente para prevenir daños y riesgos causados por cambios en los niveles del agua [5]. En este caso comparativo, la hipótesis planteada corresponde a analizar si *la variedad de contexto meteorológico y de entorno influye en el análisis de las fluctuaciones de la napa freática.* El primer caso corresponde a la zona de la cuenca mediterránea; mientras que el segundo se desarrolla en la zona del Alto Valle del Río Negro, en cooperación con expertos del Instituto Nacional de Tecnología Agropecuaria (INTA) Alto Valle.*

* Las hojas de datos funcionales (datasheets) permiten representar variedad de dominio y servicios que interactúan con la funcionalidad requerida.

* <https://www.argentina.gob.ar/inta>

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*



Fig.2. Enfoque top-down (T-VIP) para el dominio de estudio

(2) *Análisis de las hipótesis.* En esta actividad se requiere primero identificar las variables que influyen en las fluctuaciones de la napa freática. Para abordar esta identificación, el ingeniero de reuso procede al *análisis de dominios*, considerando trabajos relacionados y aportes de expertos.

De acuerdo al análisis, los factores más comunes que afectan a las fluctuaciones de la napa freática, observadas en varios estudios, son las **lluvias** y las **temperaturas** [2,5–7,11,13]. Por ejemplo, en [5] se detectó una correlación positiva entre las variables de temperatura y las recargas de la napa freática. Lo mismo ocurre en [6], donde todos los sitios muestran un aumento en la recarga con el aumento de la temperatura y también con el aumento de las precipitaciones. Respecto a los estudios que analizan las precipitaciones, en particular la **nieve** como en [7,13], se llega a la conclusión de que es un componente que forma parte de la recarga de las aguas subterráneas. También se encontró que la **humedad** puede ser un determinante de la fluctuación y se puede observar en los estudios en [8,13]. Otra variable meteorológica que se tomó en cuenta para analizar las fluctuaciones es la **radiación solar** presente en los estudios en [5,6]. En el primero se puede observar que existe una correlación entre la temperatura con la radiación solar entrante, pero se muestra mayor correlación sobre la temperatura máxima. Por último, estudios que incluyen en sus análisis el **NDVI*** y la vegetación [4,12], mostraron resultados de correlación negativa entre el NDVI y los cambios en la precipitación y la profundidad del agua subterránea, también observando que el cambio de profundidad del agua subterránea está relacionado con las condiciones atmosféricas y superficiales, como la precipitación, la temperatura, la absorción del suelo, el contenido de agua de la vegetación, la humedad del suelo y la evaporación.

* El Índice de Vegetación de Diferencia Normalizada (NDVI) es un indicador simple de biomasa fotosintéticamente activa o, en términos simples, un cálculo de la salud de la vegetación.

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*

Una taxonomía de influencias en fluctuaciones. El análisis del paso (2) identificó una serie de variables (Figura 3), que se conforma de cuatro grandes categorías: (1) Variables meteorológicas (2) Suelos (3) Tipo de vegetación y (4) Características estacionales.

Por ejemplo, la *precipitación* puede ser líquida o sólida. La primera incluye la lluvia y la llovizna; en la superficie de la tierra o en diferentes objetos, la precipitación líquida puede formarse como rocío o película líquida. La precipitación sólida incluye partículas de hielo suspendidas en la atmósfera que caen sobre la superficie terrestre. El *tipo de suelo* hace referencia a la textura del suelo, que es la proporción (en porcentaje de peso) de las partículas menores a 2 mm de diámetro (arena, arcilla y limo). Esto determina clases texturales, que se encuentran clasificadas en lo que se conoce como “triángulo textural”**. La *infiltración* se refiere al movimiento del agua en el suelo que ocurre después de la aplicación de riego o durante la lluvia, donde el agua ingresa verticalmente desde la superficie hacia el interior del suelo; y la *permeabilidad* mide la velocidad de esa infiltración. El *tipo de vegetación* se refiere a la capa de plantas que crece en un área específica, ya sea cultivada o de forma natural. La vegetación exhibe variaciones en cuanto a su tamaño, estructura y patrones estacionales, así como en los cambios que experimenta a lo largo del tiempo. Las *características estacionales* resumen aspectos que afectan el clima en períodos determinados y por diversas causas, ej. monzones o fenómeno de El Niño.

** R- 001- Guía para la determinación de textura de suelos por método organoléptico:
<http://www.prosap.gov.ar/>

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*

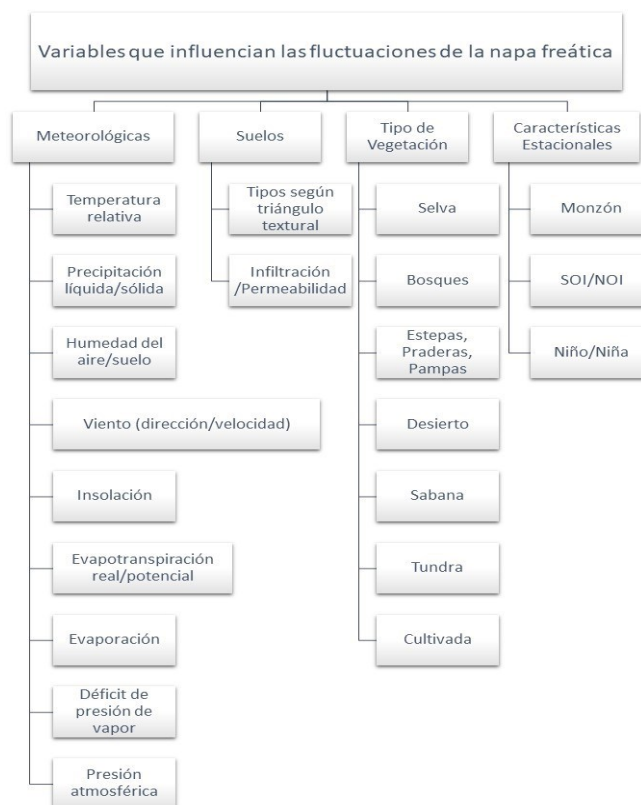


Fig.3. Taxonomía de variables relevantes en el análisis de las fluctuaciones

3. Análisis comparativo: instanciación de casos

El *Caso 1* se sitúa en la cuenca mediterránea, donde desde Junio 2017, el Instituto de Ciencias Atmosféricas y Clima del Centro Nacional de Investigaciones de Italia (ISAC-CNR*) ha tomado medidas de fluctuaciones de la napa de un acuífero en zona kárstica [3]. Las medidas fueron tomadas en dos freáticos localizados en un área suburbana, dentro del campus de la universidad de Salento, con vegetación local mixta, incluyendo arbustos de tipo mediterráneo, pinos y olivos. El *Caso 2*, se situó en el norte de la Patagonia, en el Alto Valle del Río Negro cercano a la zona de la ciudad de Villa Regina, donde el INTA Alto Valle registra las fluctuaciones de la napa freática desde el año 2010. Las medidas corresponden a dos freáticos situados en tipos de suelos y vegetación diferentes.

* <https://www.isac.cnr.it/en/node/14864>

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*

3.1 T-VIP paso a paso: ¿los datos lo corroboran?

(3) *Flujo de procesamiento.* Para cada actividad, se muestra una síntesis de lo realizado, en ambos casos, por el analista de datos.

Ingestión de datos

- Caso 1: los datos provistos en el repositorio corresponden a mediciones de nivel freático, variables meteorológicas (temperatura, precipitación, evapotranspiración, humedad, etc.), variables de infiltración, tipo de suelo y vegetación, todas obtenidas de fuentes no estructuradas.
- Caso 2: los datos provistos por el INTA Alto Valle para trabajar provinieron de diversas fuentes (datasets estructurados, no estructurados, semi - estructurados, espaciales vector y espaciales raster). La Figura 4 muestra parte de esos datos, tomando como punto de variación a los tipos y como variaciones a los lotes de datos en sí mismos.

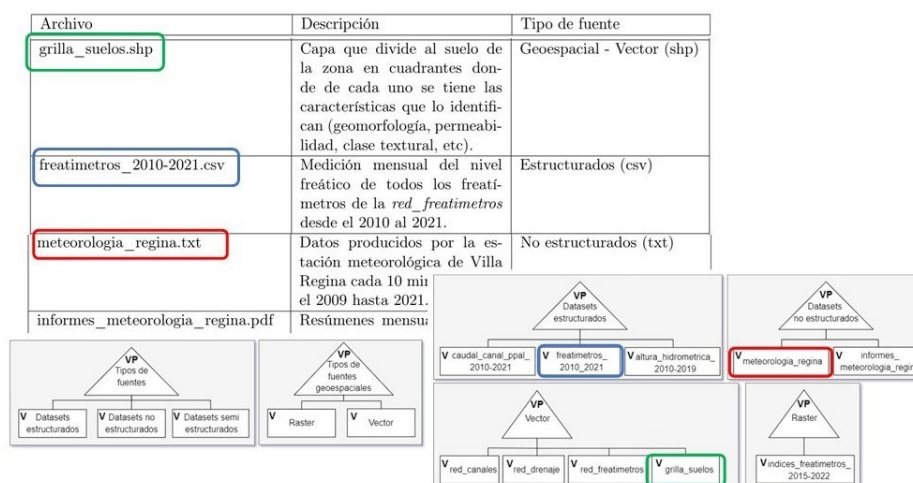


Fig.4. Variedad de las fuentes - Caso 2

Preparación de los datos. En ambos casos, la preparación consistió en conversión de datos no estructurados (pdf, txt) a formato csv. También en la identificación de campos nulos y su reemplazo analizando distintas técnicas; y en la reunión de fuentes de datos de distintos períodos de mediciones. Después del preprocesamiento, se analizó la correspondencia de variables disponibles con respecto a la taxonomía. Como resultado, ambos casos recolectaron datos de las variables meteorológicas, excepto por *humedad del suelo* (recolectado en el caso 1 pero no en el caso 2) y *radiación* (recolectado en el caso 2 pero no en el caso 1), lo que implica la detección de variedad de contenido entre ambos casos.

Análisis de los datos. La correlación, analizada con el coeficiente de Pearson para ambos casos, entre datos meteorológicos y niveles freáticos mostró (correlaciones en valor absoluto mayores a 0,5, Figura 5):

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*

- Caso 1: En ambos pozos (FW y BW), las variables que muestran mayor influencia en las variaciones de la napa son: *precipitación* (FW, 0,576411; BW, 0,513294) y *humedad del aire* (FW, 0,646698; BW, 0,668482).
- Caso 2: Se detectaron discrepancias entre ambos frentímetros. Mientras que en el frentímetro 61190 la *temperatura* (-0,641544) y la *radiación* (-0,695296) se muestran como las mayores influencias; en el frentímetro 60610 no se observa ninguna correlación significativa.

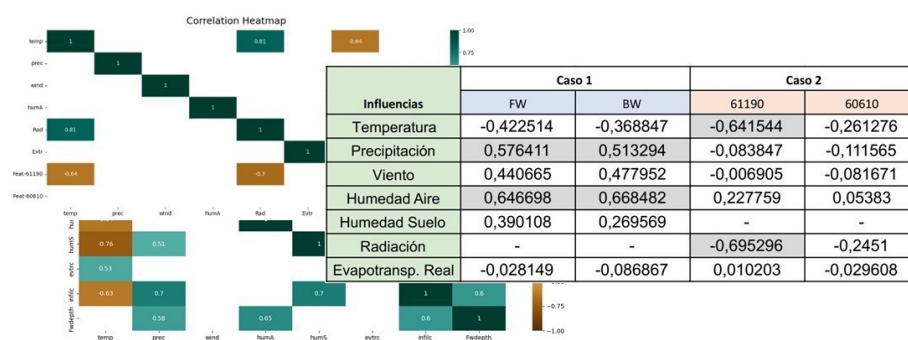


Fig.5. Correlación meteorología-nivel de la napa freática

(4) *Visualización de resultados y conclusiones del caso.* Este primer análisis muestra que las variables meteorológicas influyen las variaciones en la napa freática de distintas maneras. Viendo la taxonomía, existen otras *variables de contexto* que podrían analizarse para dar respuestas a estas discrepancias. Por ejemplo, el Caso 1 corresponde a una zona boscosa del mediterráneo, con una precipitación media anual de 1,67 mm (2017-2022), con suelos de tipo fA (franco arenoso), permeabilidad muy rápida y pozos (FW,BW) ubicados en cercanía de zona urbana (parte superior de la Figura 6). La opinión experta es que es razonable esperar que en ese contexto, las variables meteorológicas más significativas sean las indicadas (el nivel de precipitaciones sumado a la vegetación hace que la humedad del aire sea relevante). Sin embargo, el Caso 2 corresponde a dos áreas de contextos diferentes: el frentímetro 61190 se ubica en la proximidad de una zona de meseta con suelos de tipo fl/f a fl (francolimoso/franco a francolimoso), con permeabilidad muy lenta y vegetación de tipo cultivos hortalizas; mientras que el 60610 se ubica próximo al Río Negro con suelos de tipo fA-Af/A-Af (francoarenoso a arenofranco/arenoso a areno franco) con permeabilidad muy rápida y vegetación de tipo pradera (parte inferior de la Figura 6). Nuevamente, el análisis con expertos revela que esas características pueden ser determinantes: el tipo de suelo y la zona cercana a la meseta hacen que las variaciones de la napa freática sean más dependientes de la recepción de agua en forma de precipitaciones y de su evaporación por radiación, aunque la humedad del aire no sea tan preponderante (precipitación media anual 0,67mm); mientras que la zona lindante al río, con un suelo permeable, registra variaciones en la napa de menor relevancia debido las precipitaciones y radiación, ya que la cercanía del cuerpo de agua es determinante. Finalmente, los resultados muestran que la hipótesis se corrobora a través de los datos; pero se puede ir más allá, sugiriendo cambios en el modelo de caracterización; por ejemplo, podría

Identificación Top-Down de Variedad de Contexto: Un Caso de Estudio en Fluctuaciones de la Napa Freática*

agregarse la característica *proximidad de un cuerpo de agua*, para identificar el caso de zonas influenciadas por este contexto. Por último, los resultados almacenados en CoVaMaT permitirán la instanciación de futuros casos, reusando experiencias previas lo más cercanas a su contexto, para priorizar variables y métodos de análisis.

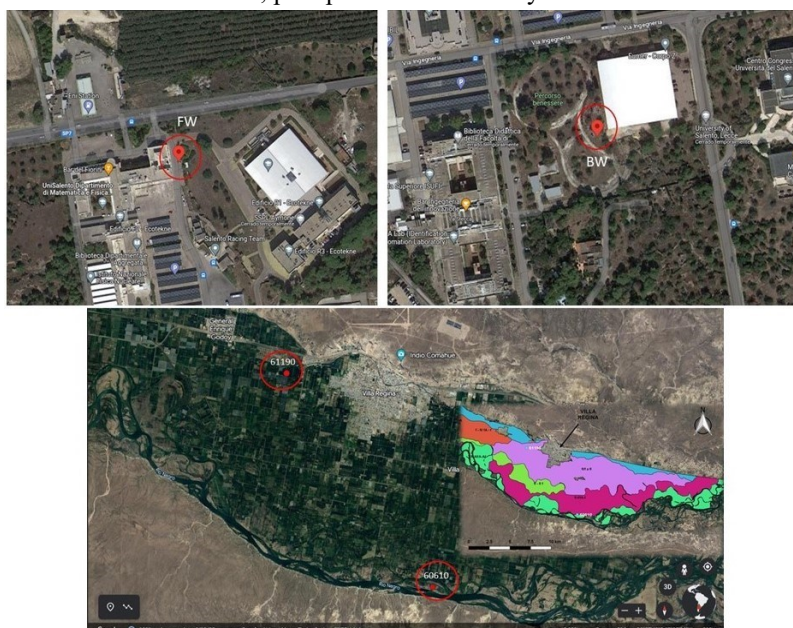


Fig.6. Variedad de contexto en ambos casos

4. Conclusiones y Trabajo Futuro

En este artículo hemos presentado el enfoque top down (T-VIP) para identificación de variedad en sistemas big data ejemplificado mediante un caso de estudio, donde la variedad de contexto es determinante para entender la relevancia de las características que definen las variables del caso de análisis. Actualmente, la investigación continúa incorporando nuevos casos y/o técnicas de análisis, a partir de la red de más de 50 freatómetros instalada en el Alto Valle. Esto permitirá seguir validando la taxonomía propuesta, como extender el modelo para evaluar el grado de reuso en cada situación.

References

1. Abawajy, J.: Comprehensive analysis of big data variety landscape. International Journal of Parallel, Emergent and Distributed Systems **30**(1), 5–14 (2015). <https://doi.org/10.1080/17445760.2014.925548>
2. Basant Yadav, Pankaj Kumar Gupta, N.P., Himanshu, S.K.: Ensemble modelling framework for groundwater level prediction in urban areas of india **712** (135539) (2020). <https://doi.org/10.1016/j.scitotenv.2019.135539>

3. Delle Rose, M., Martano, P.: Datasets of groundwater level and surface waterbudget in a central mediterranean site (21 june 2017ndash; 1 october 2022). *Data* **8**(2) (2023). <https://doi.org/10.3390/data8020038>
4. Ge Song, Jin-ting Huang, B.h.N.J.w.W.L.Z.: Effects of groundwater level onvegetation in the arid area of western china. *China Geology* **3**, 527535 (2021). <https://doi.org/10.31035/cg2021062>
5. Gonzalez, R.Q.: Prediction of groundwater levels in a changing climate. Masterthesis, MSc Geoinformatics (2021)
6. J. L. McCallum, R. S. Crosbie, G.R.W.W.R.D.: Impacts of climate change ongroundwater in australia: a sensitivity analysis of recharge. *Hydrogeology* **18**, 1625–1638 (2010). <https://doi.org/10.1007/s10040-010-0624-y>
7. L.K. Mancewicz, L. Davisson, S.W.E.B.S.P., Tyler, S.: Impacts of climate changeon groundwater availability and spring flows: Observations from the highly productive medicine lake highlands/fall river springs aquifer system. *AMERICAN WATER RESOURCES ASSOCIATION* **57**(6), 1021–1036 (2021)
8. Mahnoosh Moghaddasi, Mehdi Mohammadi Ghaleni, M.A., Salari, S.: Potentialimpacts of climate change on groundwater levels in golpayegan plain, iran. *Water Supply* (2022)
9. Osycka, L., Buccella, A., Cechich, N.A.: Data variety modeling: A case of contextual diversity identification from a bottom-up perspective. In: 27th Argentine Congress, CACIC 2021, Salta, Argentina, October 4-8, 2021, Revised Selected Papers. *Communications in Computer and Information Science* 1584. pp. 124–138. Springer (2022)
10. Osycka, L., Cechich, A., Buccella, A., Montenegro, A., Muñoz, A.: Covamat: Functionality for variety reuse through a supporting tool. In: XI JCC-BD&ET (2023)
11. Safieh Javadinejad, Rebwar Dara, F.J.: Modelling groundwater level fluctuation in an indian coastal aquifer. *Water SA* **46**(4), 665–671 (2020). <https://doi.org/https://doi.org/10.17159/wsa/2020.v46.i4.9081>
12. Shwan Seeyan, Broder Merkel, R.A.: Investigation of the relationship between groundwater level fluctuation and vegetation cover by using ndvi for shaqlawa basin, kurdistan region – iraq. *Geography and Geology* **6**(3), 187–202 (2014). <https://doi.org/10.5539/jgg.v6n3p187>
13. Zhuoheng Chen, Stephen E. Grasby, K.G.O.: Relation between climate variabilityand groundwater levels in the upper carbonate aquifer, southern manitoba, canada. *Hydrology* **290**, 43–62 (2003). <https://doi.org/10.1016/j.jhydrol.2003.11.02>