ADDRESSING THE ELEPHANT IN THE UNDERGROUND: AN ARGUMENT FOR THE INTEGRATION OF HETEROGENEOUS DATA SOURCES FOR RECONCILIATION OF SUBSURFACE UTILITY DATA

L. H. Hansen [1], R. van Son [2], A. Wieser [3], E. Kjems [1]

[1] Department of the Built Environment, Aalborg University, Denmark – (lhha, kjems)@build.aau.dk
[2] Singapore-ETH Centre, Singapore - vanson@arch.ethz.ch
[3] Institute of Geodesy and Photogrammetry, ETH Zürich –andreas.wieser@geod.baug.ethz.ch

KEY WORDS: Underground infrastructure modelling, geospatial data integration, data fusion, data quality.

ABSTRACT:

In this paper we address the issue of unreliable subsurface utility information. Data on subsurface utilities are often positionally inaccurate, not up to date, and incomplete, leading to increased uncertainty, costs, and delays incurred in underground-related projects. Despite opportunities for improvement, the quality of legacy data remains unaddressed. We address the legacy data issue by making an argument for an approach towards subsurface utility data reconciliation that relies on the integration of heterogeneous data sources. These data sources can be collected at opportunities that occur throughout the life cycle of subsurface utilities and include as-built GIS records, GPR scans, and open excavation 3D scans. By integrating legacy data with newly captured data sources, it is possible to verify, (re)classify and update the data and improve it for future use. To demonstrate the potential of an integration-driven data reconciliation approach, we present real-world use cases from Denmark and Singapore. From these cases, challenges towards implementation of the approach were identified that include a lack of technological readiness, a lack of incentive to capture and share the data, increased cost, and data sharing concerns. Future research should investigate in detail how various data sources lead to improved data quality, develop a data model that brings together all necessary data sources for integration, and a framework for governance and master data management to ensure roles and responsibilities can be feasibly enacted.

1. THE NEED FOR RELIABLE INFORMATION ON SUBSURFACE UTILITIES

Driven by a persistent and growing need to develop infrastructure above and below the surface, planners, engineers and contractors rely on information on the presence and location of unseen subsurface utilities. However, much of the available information is positionally inaccurate, not up to date, and incomplete. Reasons for this include but are not limited to a previous lack of or use of outdated survey practices, previous information representations utilising relative positions and schematic drawings, data conversion and digitisation introducing quality loss, and data quality requirements that increase over time such as the need to capture locations in full 3D. Also, the degree of quality is typically unknown, leading to increased uncertainty, costs, and delays incurred by infrastructure projects due to the need for verification.

Programs and platforms such as the Danish Register of Underground Cable Owners (LER) in Denmark (SDFE, 2021), the Cables and Pipes Information Centre (KLIC) in The Netherlands (Kadaster, 2021), and the National Underground Asset Register in the United Kingdom (Geospatial Commission, 2020) have been established to make data on subsurface utilities available in a standardised, digital format, addressing data availability and uniformity. However, accuracy and reliability of the provided records remain largely unaddressed. While legislative instruments may specify the required accuracy of utility records, it is unclear how compliance to such requirements is verified or how data owners can improve the accuracy of their data, in particular for legacy data representing utilities that were installed in the past.

2. LEGACY RECORDS: THE ELEPHANT IN THE UNDERGROUND

To improve the quality of available information, initiatives have been undertaken to increase the accuracy and reliability of "as-built" records of subsurface utilities which are captured at the time the utilities are installed. Standards and guidelines such as the Specifications for Utility Survey in Singapore (Singapore Land Authority, 2017) describe how utilities are recorded in absolute positions and with predefined positional accuracies. They prescribe the techniques, observation standards, or competencies and skills required to ensure that location information is captured with sufficient accuracy and the data attributes that are to be provided.

Such improvements address the recording of utilities directly after being built - typically when they are still exposed and direct or line-of-sight observations are possible - and do not cover the recording of pre-existing infrastructure, leaving legacy data quality issues unaddressed. As a consequence, unreliable information will continue to have a negative effect moving into the future. With multiple organisations working together on infrastructure development projects and - in dense urban areas in particular - multiple projects taking place in the same area over time, unreliable information will repeatedly lead to ineffective decision making, productivity loss, increased risks to the safety of workers and the operation of utility services, and, ultimately, extensive resources spent to deal with them.

Data on previously built assets above the ground such as buildings and transportation infrastructure can often be captured at an arbitrary moment in time to obtain data of the desired quality. The same principle does not apply to underground utilities that are not visible or accessible in their entirety and for most of their lifetime. Trials conducted in 2018 by the Digital Underground project in Singapore demonstrated that an one-off,

area-based mapping approach using 3D ground penetrating radar is not feasible nor economically viable for the purpose of improving the quality of comprehensive legacy records (Van Son et al., 2019). Instead, a gradual, long-term strategy capitalising on various data collection opportunities was deemed necessary.

We refer to newly captured data on previously built utilities as "as-is" data. Viable as-is data collection opportunities are centred around ongoing construction and maintenance projects where reliable information provides direct benefits to the parties involved in the project. Commonly referred to as a part of Subsurface Utility Engineering (SUE) practices (Zembillas & Scott, 2010), survey methods ranging from above surface observations to non-destructive surveys based on geophysics and to trial hole excavation to locate, verify, and map existing utilities. While it can be argued that such practices establish a degree of data quality improvement in support of specific projects or tasks, results from these surveys are not sustained or sustainable. Often, they are not shared or stored beyond the scope of individual projects or organisations and may not be available in a georeferenced digital machine readable form to support future use.

This loss of information between individual projects and organisations is comparable to how building design information is lost between project phases due to handover requirements in a conventional design-bid-build paper-based process. At each point between project phases, all accumulated information is downgraded into paper drawings and a laborious recreation process by the next project phase team is needed to bring it up to digital form. As a solution, many in the AEC industry are now using a digital-only collaborative Building Information Modelling (BIM) delivery process (Eastman et al., 2011). Instead of using paper drawings as records, BIM relies on digital building information models where the accumulated building information is stored, updated, maintained and exchanged between designers, engineers, stakeholders, and others (Borrmann et al. 2018). A BIM-inspired approach could potentially help sustain a higher degree of reliability of subsurface utility information across underground-related projects for any given utility asset. Figure 1 illustrates how consolidating quality improvements leads to a (more rapid) increase of information quality.
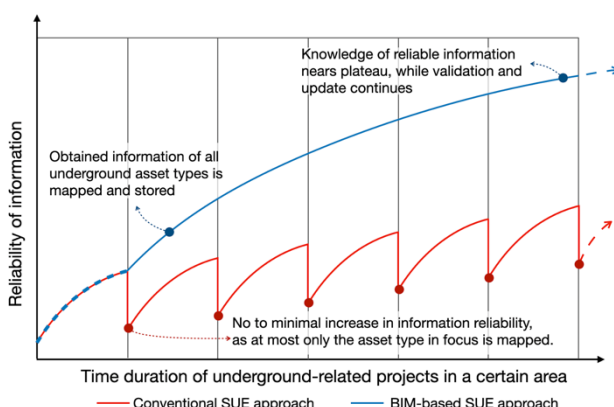

Figure 1. Loss of subsurface utility information quality caused by SUE results not being mapped and stored.

In summary, legacy data is not sustainably improved, repeatedly resulting in negative outcomes. The **purpose of this paper** is to pose the hypothesis that gradual reconciliation of legacy utility data is achievable and can be sustained through the integration of heterogeneous data sources collected at various opportunities. In the next section, a number of data reconciliation use cases are proposed and exemplified by real-world cases.

3. INTEGRATION OF HETEROGENEOUS DATA SOURCES: AN OPPORTUNITY FOR DATA RECONCILIATION

In this section, an argument is made for a novel approach towards subsurface utility data quality improvement that relies on the integration of heterogeneous data sources. Key motivations for this approach are (i) the necessity to change the status quo, and (ii) the use cases that such data integration would engender. These use cases include:

I. validation, control, and (re)classification of data quality and other attribute values of existing utility assets.

II. addition and inference of missing or incomplete utility asset alignments and attribute values.

III. improvement of positional accuracy by repositioning features or upgrading them from 2D to 3D.

The approach is to capitalise on data capture opportunities that occur throughout a utility asset's life cycle as shown in Figure 2. Opportunities to collect data on particular utilities may also occur when planning and executing nearby construction projects utilising SUE methods such as trial holes and non-destructive geophysical instruments.
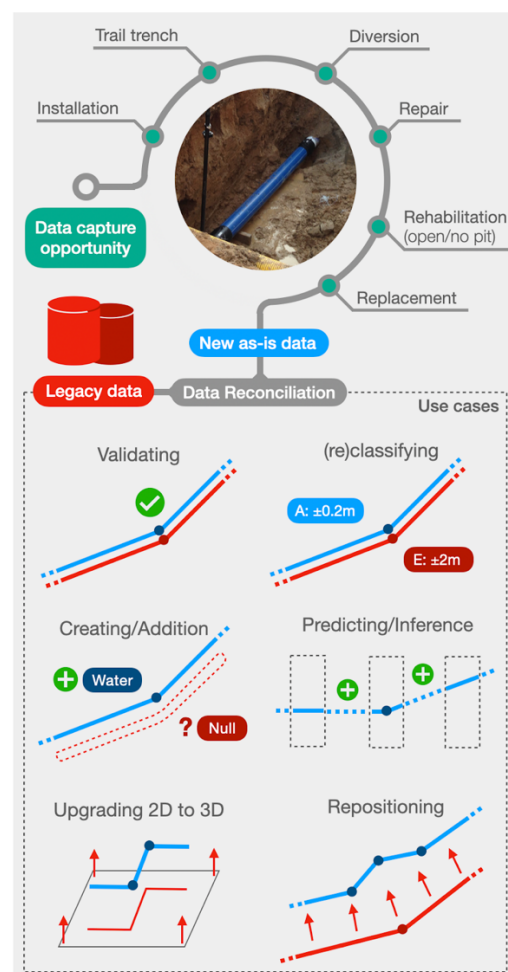

Figure 2. Data collection opportunities and data reconciliation use cases.

Throughout the mentioned data collection opportunities, a range of utility surveying and locating techniques can be utilised. In figure 3, a selection of common techniques is presented. The primary data output of these various techniques results in heterogeneous data sources that have relatively low value and an often manual processing of the data is needed to further enrich and transform the data sources into a meaningful and usable data format. In many cases this translates to 2.5D vector lines enriched with attributes, as most utility owners use a GIS-based asset management system. However, in some cases utility owners also may want to transform the primary data into a true 3D representation.
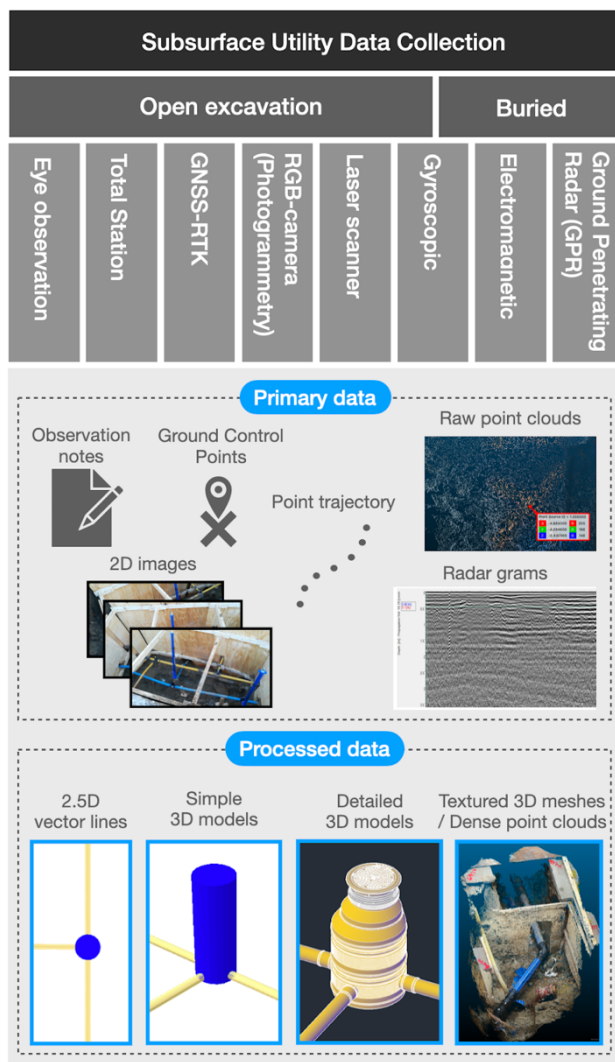

Figure 3. Common subsurface utility data collection techniques and its primary and processed data forms.

A notable category is that of "eye observation". While not a survey technique based on technology, eye observations are a potentially valuable source of data that requires a relatively low effort to capture. Platforms and programs such as KLIC and NUAR have provisions for reporting aberrant situations. In the case of KLIC for example, it is mandatory for excavating parties to report situations that differ from the situation described by the provided data. These situations are aberrant alignment, non-locatable utility, and unknown utility.

3.1 Cases of potential utility data sources

To further investigate the potential use of our data reconciliation approach, two cases from Denmark and Singapore are presented with a focus on how newly captured utility data compares to legacy records.

3.1.1 3D capture during utility replacement:
3D capture methods have become increasingly popular because of the technology improvement in laser scanners and democratisation of photogrammetry solutions. Over the past two years, two water utility companies in Denmark have tested a smartphone-based photogrammetry service as an as-built 3D documentation method during open excavation replacement of water pipes (Hansen et al., 2020a). The 3D capture solution benefits the utility companies by providing visually realistic dense point clouds of their installed water utility assets. The 3D model supports use cases such as (i) quality assurance of the agreed as-built work, (ii) visual feature extraction for completing the registration process in GIS for instance by identifying component type of the installed pipes and (iii) planning of future utility work at the same location of already 3D captured excavation holes.

A more unexpected benefit that was discovered by the utility companies was the included data capture of parts of other utilities placed near the installed water pipes. Having access to the location of these nearby utilities is expected to be highly valuable in the future when revisiting the same area as many of the other utilities were missing or wrongly positioned in the utility map records provided by the other utility owners. An example of this is visually illustrated in figure 4. Based on the utility companies experience this was a common scenario as "soft" cable records are often lacking accuracy and completeness (Hansen et al., 2020b).
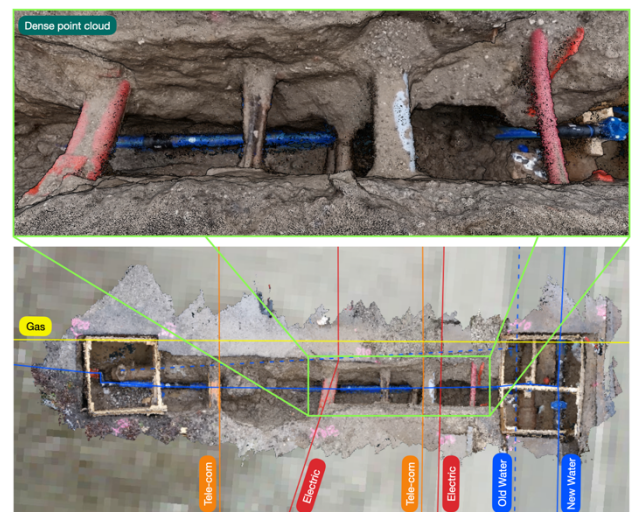

Figure 4. Comparison between existing GIS utility records and as-is excavation hole point cloud seen from above (bottom) and a zoomed in 3D view of a section of the point cloud (top).

Another common example is shown in figure 5. Besides some missing utilities it is evident that existing records are lacking completeness and detail. For example, map records do not show how many cables are located on a given utility vector line. In the point cloud, four more cables are visible compared to the utility line extracted from the map records, making the total area of occupied space larger than anticipated.
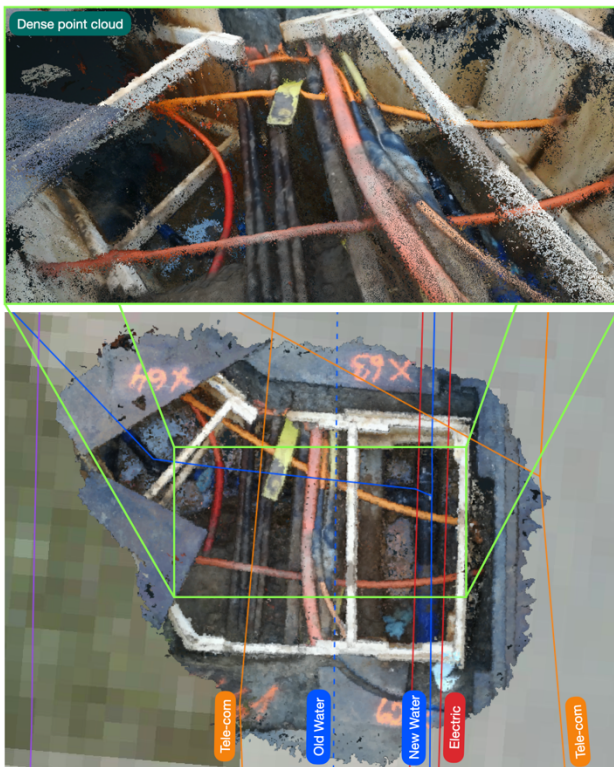
Figure 5. A similar comparison between existing GIS utility records and as-is excavation hole point cloud seen from above. The orange tele-com lines lack completeness.

For now, the 3D capture models are only used for internal use within the respective utility companies. However, the utility companies hope to potentially exchange their point cloud data with other neighbouring utility owners as 3D capturing solutions become more widespread.

3.1.2 3D ground penetrating radar data capture of large areas: Ground penetrating radar (GPR) is a non-destructive technique that can be used to detect and locate subsurface utilities using electromagnetic waves that are sent into the ground. Non-destructive techniques such as GPR can reduce and potentially even remove the need for techniques that rely on direct access or line of sight for mapping previously built utilities, reducing disruptions and nuisances, risk, and cost that come with excavations, in particular on public roads.

While a case study in Singapore in 2018 concluded that a one-off area-based mapping approach utilising a 3D or multichannel GPR is not feasible nor economically viable (Van Son et al., 2019), inspection of the data and the case study results shows that valuable information on underground conditions was obtained and that there were significant discrepancies between the detected utilities which were extracted as 2.5D points and lines and the available GIS information on existing utilities.

Notable observations from the case study were that many GIS records did not match their counterparts mapped from the GPR data with sufficient accuracy and that it was not possible to confidently match all GIS records with their GPR counterparts and vice versa. Moreover, legacy GIS records were available in 2D only, lacking elevation information that could be obtained from GPR.
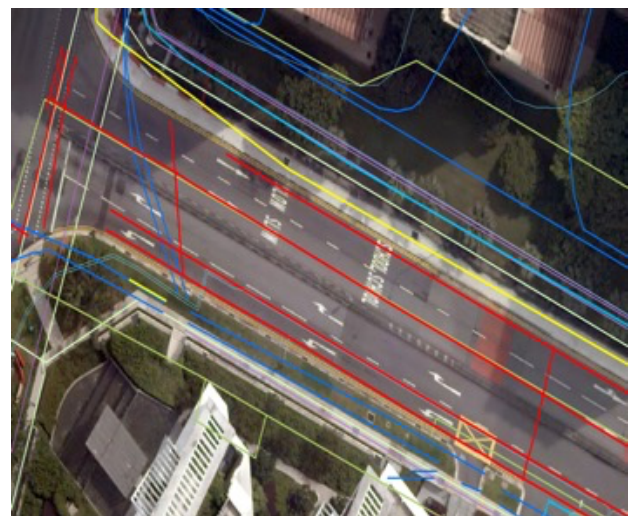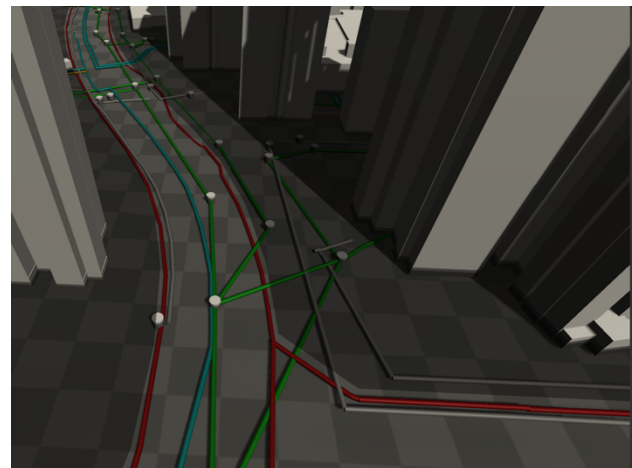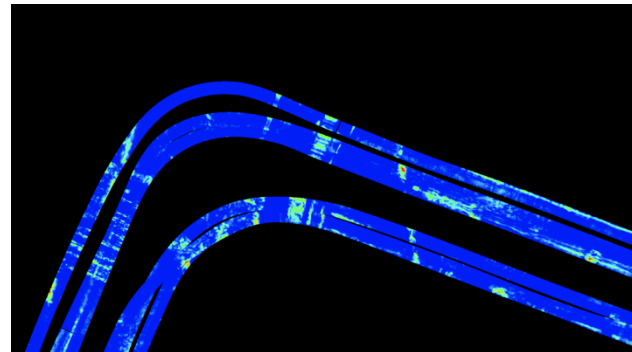

Figure 6. From top to bottom: (i) top-down view of primary GPR data, (ii) 3D render of extracted results, and (iii) comparison between utilities from GPR scans (red) and available GIS records

While additional observations would be necessary to confidently link GIS records and GPR vectors, the results demonstrate both the need for legacy data reconciliation in Singapore and the potential use cases that could be supported, which include validation, upgrading from 2D to 3D, and repositioning.

4. CHALLENGES IDENTIFIED IN PRESENTED CASES

From the presented cases, a number of challenges could be identified that range from social to financial to technical ones. The first is technological readiness. Adoption of state of the art survey techniques such as those based on photogrammetry and geophysics was observed to be low in Denmark and Singapore and surveyors would usually opt for conventional, direct measurement techniques instead. However, the smartphone-based Reality Capture solution used by the two utility companies in the Danish case was concluded to be a feasible surveying solution indicating an encouraging sign of achieving higher technology readiness (Hansen et al. 2020a). Moreover, asset owners' data management systems are often not yet able to ingest, store, share, or use data captured in complex and rich 3D representations as well as data with varying degrees of quality and fidelity.

The second challenge is a possible lack of incentive to capture the necessary data or improve data quality. In many jurisdictions around the world, utility companies are not liable for the quality of information that they provide. Furthermore, the example cases show that there are opportunities to survey types of utilities that do not belong to or are of interest to the companies mandating or performing the work. It would be questionable to assume that such companies would invest effort and resources in capturing, improving and sharing such data. This links closely to the third challenge which is that of cost. Performing the necessary data capture during suitable opportunities and upgrading data management systems to handle new data sources requires a significant financial investment that is unlikely to yield a return in the short term. And fourth, it may not be desirable to make information on certain utilities known between parties due to security and business concerns.

5. CONCLUSSION AND FUTURE WORK

The integration of heterogeneous data sources captured at various opportunities during the life cycle of subsurface utilities could be used to improve data quality and reconcile legacy data. Example cases show that data captured using techniques such as photogrammetry and ground penetrating radar could be used for various quality improvement use cases.

To achieve the objective of gradual and sustained improvement of data quality, critical challenges need to be overcome. We propose that further research focuses on three key elements that together could form the basis of a robust framework for the reconciliation and improvement of subsurface utility data.

5.1 Investigate specific data quality improvement scenarios

First, future research should develop a comprehensive overview of relevant data sources for quality improvement. It should investigate what heterogeneous data sources with different quality (e.g., accuracy, reliability, resolution) can contribute to data quality improvement and in what way (e.g., validation, position accuracy improvement). Besides the techniques demonstrated by the examples in this paper, the research could consider eye observations which do not result in geometry or location information but rather information about it (e.g., on aberrant alignments, or confirmations of correct alignments). From a pragmatic perspective, it is recommended to focus on data capture opportunities that are already occurring but are not yet utilised to their full potential. Open trench excavations - both for when new utilities are installed and existing utilities are partially exposed and trial holes to verify existing utilities - are logical starting points as they are typically unavoidable.

5.2 Development of a data model

A data model needs to be developed that meets a number of requirements in order to facilitate data integration and data quality improvement. First, the data model needs to be able to integrate and connect various data sources, ranging from legacy data to newly captured data sources. To enable a degree of automation for quality control and quality improvement, Integration should be established through more than georeferencing alone, for example by establishing a common and persistent reference to physical utility assets or structures. Second, the data model needs to be able to support a range of data capture techniques and data types. For example, while legacy data may be available as 2D GIS or CAD files, newly captured data could represent utilities as 2.5D or true 3D geometry. It is also important that both primary and processed data can be integrated and stored, as primary data sources could serve as valuable sources for future data quality improvement. Third, the data model needs to clearly define data quality and its descriptors (e.g., accuracy, completeness, consistency) in order to measurably assess and improve quality.

It is recommended to further build upon data models for subsurface utilities that are designed to integrate various data sources such as the MUDDI model (Lieberman, 2019) and the Singapore Underground Utility Data Model (Yan et al., 2021).

5.3 Development of a framework for governance and master data management

There needs to be a clear definition of the roles and responsibilities of all stakeholders required for data reconciliation. National utility asset information exchange systems such as those in Denmark, The Netherlands, and the United Kingdom are all organised as variations of decentralised "registry" architectures where utility owners manage and maintain data pertaining to their own assets and make their data available to users through a common portal (Figure 7 top part). In such cases, the responsibility for data quality improvement is assumed to be primarily with the utility owner. For such cases, the effectiveness of legislation needs to be assessed. Relevant examples include the direct mandate of quality improvement in France (Zeiss, 2021) where utility owners are responsible for improving data quality when they are not able to provide them to requesting entities at the indicated quality level, output-oriented accuracy requirements for provided data (e.g., ±1m horizontal accuracy for utility data in The Netherlands), or indirect incentives imposed by regulators such as on asset resilience which is assumed to be affected by unreliable information resulting in excavation damages (OfWat , 2019).

However, identified challenges such a lack of technological readiness and incentive among individual utility owners may result in a siloed and ineffective approach to data reconciliation. Instead, a more centralised approach where data is stored and improved in a single, dedicated system could be explored as well (Figure 7 bottom part). Such a system would collect survey results directly from relevant opportunities such as construction projects, reconcile (legacy) data stored inside, and provide the updated and improved results to the individual utility owners and other beneficiaries.
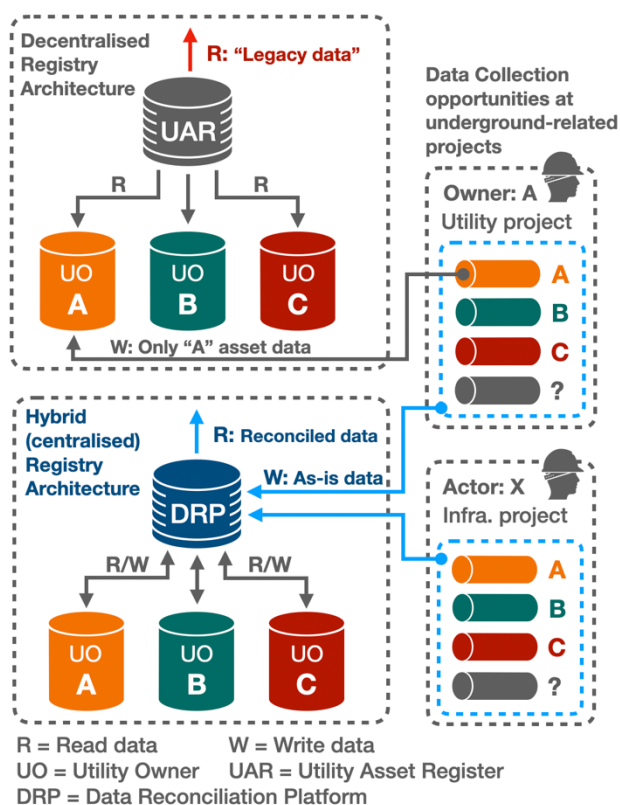
Figure 7. Traditional decentralised-based Utility Asset Register (top part) compared to a centralised-inspired approach (bottom part) that integrates a Data Reconciliation Platform to reconcile sourced Utility Owner data with new captured as-is data.

REFERENCES

Borrmann, A., König, M., Koch, C., & Beetz, J. (2018). Building information modeling: Why? what? how?. In Building information modeling (pp. 1-24). Springer, Cham.

Eastman, C. M., Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors. John Wiley & Sons.

Geospatial Commission, 2020. National Underground Asset Register Project Update. https://www.gov.uk/government/publications/national-underground-asset-register-project-update, Accessed July 2021.

Hansen, L. H., Wyke, S. C. S., Mandahl Pedersen, T., & Kjems, E. (2020a). New Technologies for the Prevention of Excavation Damages: Using Reality Capture and Augmented Reality. *Geoforum Perspektiv*, *19*(37), 49-54. https://doi.org/10.5278/ojs.perspektiv.v19i37.6278

Hansen, L. H., Wyke, S., & Kjems, E. (2020b). Combining Reality Capture and Augmented Reality to Visualise Subsurface Utilities in the Field. Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC), 703–710. doi.org/10.22260/isarc2020/0098

Kadaster, 2021. KLIC Kabels en Leidingen Informatie Centrum, https://www.kadaster.nl/zakelijk/registraties/landelijke-voorzieningen/klic, Accessed July 2021.

Lieberman, J. (Ed.), 2019. Model for Underground Data Definition and Integration (MUDDI) Engineering Report, http://www.opengis.net/doc/PER/MUDDI-ER, Accessed July 2021.

Ofwat (2019). PR19 Final Determinations. Overall stretch on costs, outcomes and cost of capital policy appendix, https://www.ofwat.gov.uk/regulated-companies/price-review/2019-price-review/final-determinations/, Accessed July 2021.

SDFE 2021. The Danish Register of Underground Cable Owners (LER). Danish Agency for Data Supply and Efficiency (SDFE), https://ler.dk/Portal/P.10.English.aspx, Accessed May 2021.

Singapore Land Authority, 2017. Standard and Specifications for Utility Survey in Singapore. Singapore Land Authority https://www.sla.gov.sg/Press-Room/Circulars/Land-Survey, August 2017.

Van Son, R., Jaw, S.W., Wieser, A., 2019. A data capture framework for improving the quality of subsurface utility information. *SPRS - Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci.* 97–104. https://doi.org/10.5194/isprs-archives-XLII-4-W15-97-2019

Yan J, Van Son, R., and Soon, K. H. (2021). From underground utility survey to land administration: An underground utility 3D data model. Land Use Policy 102, 105267. https://doi.org/10.1016/j.landusepol.2020.105267

Zembillas, N., M., and Scott, P., C., (2010), The Evolution of Subsurface Utility Engineering: From concept to Industry Standard, North American Society for Trenchless Technology (NASTT), No Dig Show 2010, Chicago, Illinois, May 2, 2010.

Zeiss, G. (2021). France's DT DICT system effective in reducing damage to underground utilities during construction, https://geospatial.blogs.com/geospatial/2021/07/french-dt-dict-system-effective-in-reducing-damage-to-underground-utilities-during-construction.html, Accessed July 2021.

(PDF) Exploring city digital twins as policy tools: A task-based approach to generating synthetic data on urban mobility. Available from: https://www.researchgate.net/publication/353288963_Exploring_city_digital_twins_as_policy_tools_A_task-based_approach_to_generating_synthetic_data_on_urban_mobility [accessed Aug 16 2021].