Consolidating feedbacks and expertise of Digital Twins of Territories' engineers in nation-wide frameworks

Théo Szanto, Mathieu Brédif, Bénédicte Bucher

Univ Gustave Eiffel, Géodata Paris, IGN, LASTIG, F-77454 Marne-la-Vallée, France - firstname.lastname@ign.fr

Keywords: Digital Twin of Territory, Metadata, User Feedback, Minecraft, Data integration

Abstract

Digital Twins of Territories (DTTs) are increasingly adopted by municipalities to support ecological transition, crisis resilience, and participatory decision-making. Designing a DTT that fits local needs requires engineers to combine multiple areas of expertise (data discovery, integration, modeling, visualization, and stakeholder interaction) while working with heterogeneous geospatial datasets of varying quality. Nation-wide DTT frameworks aim to assist these efforts, yet they currently lack mechanisms to consolidate the expertise produced during local DTT developments. This paper introduces `dttrecipe`, a model designed to capture, structure, and share DTT engineers' feedback and decision-making processes. Building on the `prov`, `wfdesc` and `wfprov` ontologies, and inspired by the OGC Geospatial User Feedback standard, `dttrecipe` formalizes the description of territorial stakes, data workflows, encountered problems, and the rationale behind design choices. It supports both complete and partial workflow descriptions, encouraging collaboration, reproducibility, and cross-territorial knowledge reuse. The model is qualitatively evaluated via a case study focused on bicycle-mobility planning and citizen engagement in a rural city. The resulting recipe highlights recurrent categories of DTT engineering challenges, including data discoverability and usability issues, multi-source misalignment, documentation accessibility, and limited local expertise. Explicit documentation of these challenges shows how engineers' often implicit expertise can be converted into reusable knowledge for other territories facing similar constraints. The work shows that structured documentation of DTT engineering practices can strengthen national DTT frameworks by improving interoperability and enabling efficient knowledge transfer. Future work will address querying mechanisms and evaluate the reuse of shared recipes at scale.

1. Introduction

Municipalities face unprecedented challenges to engage their territories in ecological transition, and to become more resilient to crises. They must increase energy efficiency, reduce soil sealing, develop more inclusive communities, adapt to more frequent climatic events and cannot afford trial-and-error experiments because of the urgent character of transition and of a lack of resources. To decide the right trajectory for their territory, they need to set a dialogue between many actors of these territories, like inhabitants, private sector, associations, and also with scientific communities, and other territories. In that context, Digital Twins, applied to cities or national scale, emerges as a solution to better assess the different priorities of a territory, to embrace its multiple dynamics, elaborate experiments, compare results, and make enlightened and shared decisions with all concerned actors (Ellul et al., 2024, Béler et al., 2024).

Originally from industrial domain, a Digital Twin (DT) is ”a realistic digital representation of real assets, processes and systems” with ”some sort of bi-directional connection to the real twin” (Callcut et al., 2021). Among different definitions in the literature, three components can be identified: a physical reality, a virtual representation, and information exchange between the two (VanDerHorn and Mahadevan, 2021). This flow of information between the reality and the twin does not need to be real-time, and may or may not involve humans in the process (Agrawal et al., 2023). DT are now studied to apply digital technologies, in particular simulation, to more systemic issues at larger scale (Metcalfe et al., 2024) to meet the unprecedented challenges mentioned in this introduction: to support automated simulations of phenomena, like pandemic, flood, but also mobility, energy consumption and to evaluate the impacts of decisions and planning scenarios. This paper focuses on the application of the DT paradigm to a territory, which is defined as a geographic area belonging to or under the jurisdiction of a governmental authority.

There is no unique form of Digital Twin of a Territory (DTT). For the DTT to be adapted to the decisions it will inform, local stakeholders have to clarify the territory stakes and priorities. Secondly, to make the most of opportunities offered by data and technologies, engineers must specify the solution that is adapted to the described stakes and that is feasible in the specific context. DTT engineering requires expertise across multiple domains (Xavier et al., 2025), including dialoguing with stakeholders to understand their priorities, identifying the form of DTT that is adapted to these priorities and to the local context, selecting the most appropriate data available on that territory (Bogdanović et al., 2015), addressing technical difficulties that inevitably arise from mixing data from multiple sources (Breunig et al., 2020), and selecting an environment suitable for the end users (Bleisch, 2012). DTT engineer may need to create 3D visualizations to engage citizens in public debates, which provide better comprehension for both specialists and general public (citizens, children...) (Onyimbi et al., 2018, Konstantinidou et al., 2023). Nation-wide DTT frameworks emerge to support the development of DTT by providing core data everywhere, and to interconnect different local and thematic DTTs together based on location (Ellul et al., 2024). In such nation-wide DTT context, knowledge exchange between DTT engineers can be particularly relevant to address the different expertise gaps faced by aforementioned DTT engineers (Coetzee et al., 2020). While a given territory has specific stakes, it can get inspiration from DTT design processes from other territories with similar interests and more experience, for example with the 3D skills they may lack. Such frameworks also aim at bringing more coherence in the DTT engineering practices among different territories.

This paper addresses the following question: how to support knowledge exchange between DTT engineers to fill their gaps of expertise and assist them to design a DTT adapted to the local stakes and context expressed by the stakeholders? It first reviews related works (section 2), then proposes a new way to model DTT conception workflows (section 3) applied to a real world example (section 4). Finally, it concludes with limitations and future works (section 5).

2. Related works

Scientific communities share knowledge and experiences mainly through publication of papers in well defined processes where the peers evaluate if a paper respects specific criteria to be published as a scientific contribution for the community. Additionally, if there is any research data, its publication is encouraged in specific platforms, together with metadata to make the data discoverable and interoperable (Wilkinson et al., 2016). Practically, researchers are encouraged to add not only formal metadata but also a readme file where they explicitly share any relevant information about the context of data production that could not be expressed through the formal metadata standard. Specific metadata standards have been proposed that are adapted to the complexity of geospatial information, as the ISO 19115 standard from the International Organization for Standardization. In a DTT context yet, other communities must be considered, like the built environment, smart city. A transversal model to describe datasets from any domain is the DCAT vocabulary proposed by the World Wide Web Consortium.

User feedback is increasingly regarded as a valuable asset on information infrastructures to support users' searches for assets. They complement official descriptions of assets with additional information sometimes more relevant to users. This is referred to by User Generated Content (UGC) (Zhang et al., 2024). The OGC standard Geospatial User Feedback (GUF) proposes a model to structure feedbacks of users of geospatial data or tools so that the different feedbacks can be better consolidated and used to inform other users choices. A feedback can include textual comments, links to papers, to code snippets or related work (Zabala et al., 2021). A DTT engineer can describe his experience as a feedback about reusing data, for example national lidar data products, and reusing pieces of software and platforms. Apart from the model to express and exchange the feedback, another stake is the motivation of users to share these, as this is a key element for the success of a solution based on user generated content (Zhang et al., 2024). Besides, there is a need of a centralized solution to gather that content; as there is no single platform to access data and technologies to develop DTT, there is no obvious actor to gather and serve these feedbacks.

Models to describe workflows, like data processing workflows, or processes, that can include manual tasks, have been studied for developers to describe and reuse steps and sequences of steps that can be automatically interpreted and run, or to describe more globally how to achieve some objectives. The DTT engineering experience can be seen as a process and partly as a workflow, although it is not intended for automatic execution. The `prov` ontology can describe actors and activities, allowing a high-level modeling of a process that yielded a given object, i.e. its provenance (Moreau and Groth, 2022). Models such as `wfdesc` and `wfprov`, are derived from `prov`, to describe workflows (Belhajjame et al., 2015). `wfdesc` allows to describe workflows in-depth with each step, inputs and outputs. `wfprov` allows to track previous executions of a workflow described by `wfdesc`. A `wfdesc:Workflow` (subclass of `prov:Plan`) is defined as a set of `wfdesc:Processes`, which are the steps, connected to each other through `wfdesc:Inputs` and `wfdesc:Outputs`. It can thus represent any sequence of tasks, either manual or automated.

3. Methodology

Our philosophy is that DTT engineers can interact with a shared knowledge base to feed it with their experiences and/or find expertise about their specific scenario or problems. Having a model to represent the DTT design process is essential to structure that knowledge base and opens efficient querying of knowledge.

3.1 The `dttrecipe` model

Our work aims at providing a model, `dttrecipe` (fig. 1), for DTT engineers to document the full process of designing a DTT: the expression of stakes, each step accomplished to create the specific DTT, including how to solve obstacles. This model enables DTT engineers to share and reuse the expertise from one another, by reproducing a recipe. The formal model must support queries to allow the DTT engineer to discover relevant recipes, possibly parts of a recipe, such as a description of stakes or the usage of specific data sources among a collection of recipes. The model is designed to document everything done by the DTT engineers, as well as why it was done that way. This includes the interpretation of the stake, the actual DTT engineering, and the description of how the DTT is to be used to meet the stakes.

This model is based on the `wfdesc` and `wfprov` ontologies, making it generic enough to represent any step needed to process data. It is extended to explain why steps are necessary, and problems they solve. Problems are categorized into problem types. Such categorization is likely to help querying a knowledge base of problems. Each kind of problem should provide indication on how to document one. The set of types of problems is not limited and we will provide six to start with.

This model can encode either a complete recipe, or an abstract one. An abstract recipe is incomplete, and may not enable reproducibility of the work, as opposed to a fully described one. For a recipe to be complete, all of its steps should be explained, but not all of them need to be implemented. A manual step, if described enough, can still be reproduced. However, a recipe containing such steps will not be executable automatically, and may require a higher level of expertise to be understood.

The proposed structure of a description in our model is the `richtext`, a text written in natural language, with annotated parts, making it easy to read both for humans and machines. Its content is unconstrained, but editing can be assisted to help users use the same names when talking about the same objects, or to suggest properties. Such annotations effectively reference other resources. The text can be written in any language, thanks to the advances of machine-assisted translation tools. Description templates can also be provided, or examples from other users. We believe that this middle ground between a completely unconstrained text and fully structured data helps users express

https://doi.org/10.5194/isprs-annals-XI-1-2026-355-2026 |

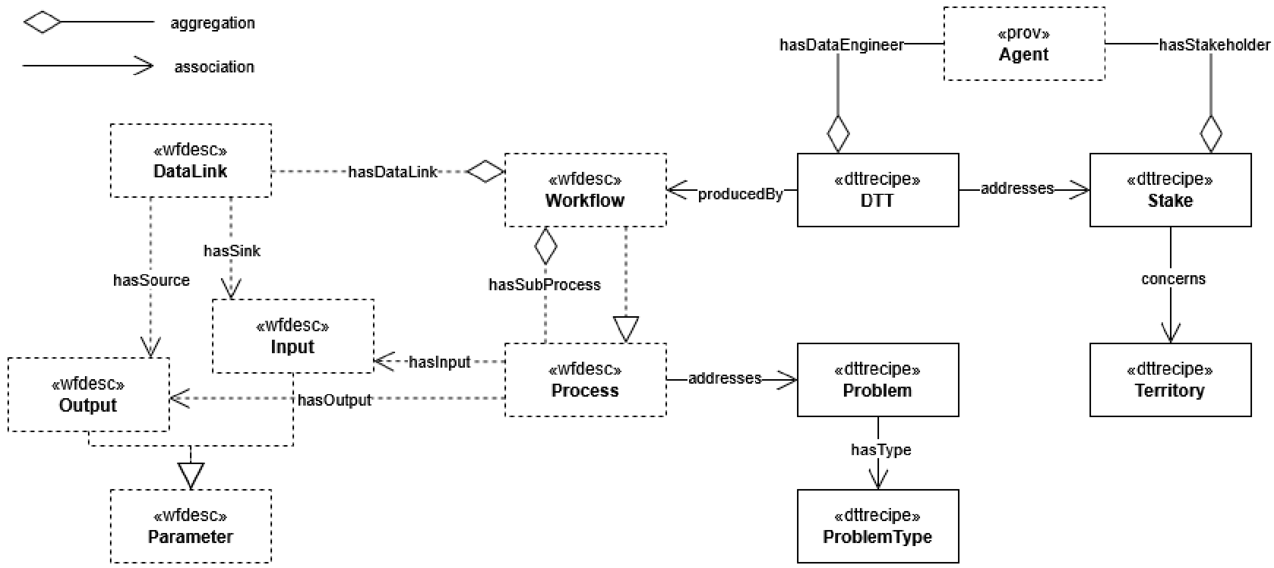

Figure 1. (UML notation) With solid lines, our model `dttrecipe`, and its relations with other models in dashed lines (`prov`, `wfdesc`)

themselves, while maintaining a basic understanding of the content by the machine.

The root element of our model is the `dttrecipe:DTT`. It represents the whole process of elaboration of a DTT, including the explicit description of the stake it addresses to the workflow to produce it. It is associated with some high-level abstract. It also references its authors, and a section about the usage, i.e. how the digital tool produced by the DTT engineers should be used within the territory to meet the stake that led to its design. The `dttrecipe:Stake` is the real world issue addressed by this `dttrecipe:DTT`. The author should describe it the most accurately possible, including all contextual information such as the project, funding, human resources involved, and their expertise level. Documenting all this information will help users find and adapt recipes to fit their situation. We do not consider it necessary to classify use cases into types, because such classifications are often subjective, and modern natural language processing techniques provide interesting results. To assist the author, some annotations (such as ”location”) could be recommended, and an automated real-time feedback system could give hints to include missing annotations. The `dttrecipe:DTT` is produced by a `wfdesc:Workflow`, which describes the work done by DTT engineers to go from the expression of the stake all the way to the results.

3.2 Writing a DTT recipe

This subsection provides guidance on how the DTT engineers should document their work to form a DTT recipe.

3.2.1 Describing the stake The very first step done by the DTT engineer is to engage in discussions with the stakeholder to understand its stake. Once it is clearly identified, it should be written down, and ideally validated by both parties. It serves as a common ground truth of the objective to reach, and may help when taking design decisions. Furthermore, it is key to later find DTT recipes designed for a similar stakeholders.

Then, the DTT engineer should explain his interpretation of that stake, justifying choices, and indicate when complementary information were given by the stakeholder. This interpretation is the translation of the real stake into DTT goals.

3.2.2 Describing the data integration A workflow is often structured like this:

1. Data provision (discovery, selection and retrieval)
2. Pre-processing
3. Main treatments
4. Post-processing
5. Export / encoding

Each kind of process plays a different role and is important to explain and justify accordingly. While we do not explicitly materialize such categorization in our model, we believe this knowledge can help the author to know what is important to document for each process of his workflow, and could be accounted for in a tool assisting him.

Data provision processes load data, sometimes reading local files, sometimes fetching remote API. When using local data, it is absolutely required to explain how to get them. With remote data, it could be interesting to give documentation about the API or catalog or data provider used, in order to help adapting the process if the remote changes the way to retrieve data. In case where a specific version of the data is needed, it must be properly indicated, even if this version is the default one at the time of writing (because it could change in the future). Whatever the data source is, the data itself should be described, so someone could find alternative data sharing the same characteristics when reusing the process. Such alternative data could also be indicated directly in the description, if known. Another important thing is to explain how to find documentation about those data, as it is not always obvious.

Pre-processing steps are particularly interesting, as they could originate from issues regarding input data. For example, a lidar point cloud may contain points classified with non-standard classes, and a process would translate those classes into standard ones, to comply with a tool accepting only standard classes. As such, having them documented is useful both for users to better understand specificity of the data, and for the data producer himself, to see how he could improve the usability of its dataset. It is likely that those steps could be reused in an entirely different context, simply when reusing the same data source. What is really important to document about those steps is their justification in the workflow, i.e. what (technical?) problem do they solve.

Main treatments depend on the goal achieved by the workflow, but are generally independent from the data sources used, by opposition to the pre-processing steps. Having them decoupled allows to replace a specific data source by another equivalent one, without having to change the whole workflow. When documenting those processes, it is particularly recommended to explain precisely what are the data requirements in order to run them, especially when those requirements are already satisfied by the input data, and thus not explicitly visible through pre-processing steps.

Post-processing steps are similar to pre-processing ones, except that they do not relate to the input data, but rather the output environment. They aim to allow the user to swap the output environment with another one, while still keeping the beginning of the workflow the same. Such modularity promotes reusability and efficient work sharing across multiple use cases.

Finally, export or encoding processes produce results appropriate to a given output environment. Those results are the ones of the workflow, in other words they are what the original author produced in order to help the stakeholder solving his use case. Thus, those are not final results, and the documentation should explain how to use them (especially aiming at a non-expert user, such as a stakeholder).

Additionally, each process should describe its outputs with enough details to understand whether they fit specific requirements, to allow users to select relevant processes without running them. Then, there should be enough details to understand if the result we got is conform (i.e. if we made no mistake when applying the process). Finally, in case we are not able to apply the process due to any reason, its outputs should be described enough to find alternative methods to obtain similar ones.

3.2.3 Describing problems During their work, DTT engineers will encounter problems related to data and processes. Such problems may vary in complexity, but always take time and efforts to solve. Documenting them as they are found allows sharing the expertise and overall reduces the global time spent on them. Once a problem is solved, it should be associated to the already documented problem, so everyone can benefit. Furthermore, problems sometimes rely on informal knowledge transmission (oral conversation between data producer and engineer) or ”hidden” (not easily discoverable/accessible) documentation to be solved, which would inadvertently prevent others from fixing them.

Six types of DTT engineering problems have been identified empirically by abstracting the set of specific obstacles encountered by the engineering of a DTT prototype in a first experiment:

Lack of expertise/knowledge is the most generic type of problems. The most important part to write about them is the initial goal, and what have been attempted to do.

Data discoverability/usability problems occur when data sources are not easily findable, accessible, or exploitable, due to the absence of referencing in catalogs, lack of metadata, or use of non-standard formats. They can be mitigated by sharing links, documentation or additional resources related to those data, to assemble a collective knowledge base around them.

Data misalignment problems can occur when integrating multiple sources of data together. This can be the case when multiple data sources describe the same real object, but both claim different, conflicting properties, and no ground truth is available. This is one of the most challenging problems to solve, and it is particularly important to share every expertise. Documentation from both data sources can help to better understand the issue. Any reconciling methods or realignment algorithm, if any, should be linked with the solution.

Data quality problems are related to a single data source. They manifest when the data contains errors, either from human mistake or malfunctioning sensor. In those cases, the impacted objects should be specified using their identifiers, and if possible, a patched version of the features could be included.

Documentation accessibility problems are a variant of data accessibility issues, but dedicated to its documentation. Sometimes, it can be hard to find, because it is not always included inside the data itself, or hard to manipulate. Huge amounts of information can be hard to navigate. Relevant links, guides, cookbooks or usage examples could be shared to address this issue.

Lack of resources is a problem that can occur during the DTT engineering, but is not always caused by the involved datasets or processes. The work may not be fully conducted due to a lack of time, funding, computational or human resources. However, solutions that are expected to work could be indicated so that someone else may try them when working in a different context, with more resources, to improve the overall results. Using massive datasets or relying on deep learning algorithms could increase the chances of encountering such issues.

3.2.4 Describing usage In the current state of this work, we did not study the usage of the DTT produced by the recipe. This will be considered in the continuation of this work. We intend to include a description of the usage in our model, which would improve reusability.

4. Results

To qualitatively evaluate our model, we applied it to a real world example. This example is taken from a funded research project to develop DTT.

4.1 The work of the DTT engineer

Territory: The territory is a small rural city in France of approximately 7000 citizens whose municipality is very prone to engage in collaborations oriented towards sustainability, and would like to further extend an initial DTT to address additional challenges. This is typical of rural cities, which makes this example a good first approach.

Agent: The DTT engineer is a researcher with data engineering skills who had to develop a first DTT prototype to meet a specific stake.

Stake: The description of the stakes was done during a face-to-face meeting. The stakeholder is a group formed by the mayor, its deputy and the technical deputy. Their city is organized around two centers of activity, separated by green fields and a highway and connected by a small country road. Walking from one to another would take approximately 10 minutes, and less than 5 minutes by bicycle. Yet, citizens prefer to take their car, because that road is not suitable for other ways of transportation, as it lacks sidewalks, road markings or any infrastructure to protect cyclists or pedestrians from cars. Besides, they wish to develop the usage of existing cycling route that leads to a leisure park outside the city and that families don’t use. Thus, the municipality now studies different solutions firstly to engage its citizens in using bikes and secondly in developing existing roads to better support bike mobility. They need help to elaborate scenarios -where to develop new cycling path- and to gather feedback from citizens who would be primarily concerned by this change, and need them to explore proposed scenarios and express their thoughts. They do not have a full-time dedicated geospatial data expert, and only a small budget allocated to this.

The following paragraphs reflect the choices made by the DTT engineer, to produce a DTT that could be used by the stakeholder. Those choices are justified and the initial objectives are stated. This allows other experts to express critical opinions on them, and ultimately could lead to improvements beneficial for everyone. Some choices can be driven by the availability of data or tools at the time of the work (November 2024).

DTT: A form of DTT solution is described hereafter that has been proposed to the stakeholder. The need to debate with citizens was identified by a key functionality and the specified DTT includes a 3D, interactive visualization, that would be easy to navigate for everyone. Interaction with 3D can be hard to achieve with professional modeling tools designed for expert users that have a steep learning curve to properly interact with them. Serious gaming can help with this. Geocraft, a project made by researchers in The Netherlands, was created using real world geospatial data, and was tested on use cases about urban area management, evaluating citizen participation (Scholten, 2017). Minecraft has proven suitable to run simulations for educational purposes, with reduced precision but enough realism to explain phenomenon (Lecordix and Kumarasamy, 2021). Furthermore, very young public can be engaged that way, which may be desirable, as the target is families, including children of all ages (de Andrade et al., 2020). The benefit of such a solution is that this environment is already used at primary and secondary education, hence all pupils could be motivated to participate and help older inhabitants. The area was limited to studied road, although representing a larger area could give visual hints about landscape, which could improve immersion and spatial awareness of citizens, used to see the reality.

Process 1: Then, the DTT engineer explored data catalog from the NMA (IGN) to find available data describing the territory. It should be noted that they had previous experience with such data, and knew which ones could be used. Considering the goal of depicting the visible part of the territory, the selected data included orthophotograph (BD ORTHO), aerial lidar point cloud (LiDAR HD), elevation data from DTM (BD ALTI), and topographic features from the national database (BD TOPO): buildings, roads, fields... All those data are available at national scale, making this representation reproducible on any territory at least from the same country.

Problem 1 (data discoverability/usability): Local data were considered and final decision was not to use them because they were not easily accessible and reusable because they were not already in geospatial format, nor provided by a single source.

Process 2: Once data were selected, the integration work began. Terrain (ground) generation was done by projecting orthophotograph onto the DTM, which did not cause any problem.

Problem 2 (lack of expertise/knowledge): Roads were then rendered initially by projecting thick lines onto the ground, with a monochrome dark gray color. This however caused an issue with bridges, which were not properly rendered. To fix this, the solution was to draw a straight line between the two ends of bridges, which were clearly recognizable thanks to their ”elevation relative to ground” attribute being strictly positive.

Problem 3 (data misalignment): The point cloud comes from aerial lidar acquisition and is then classified using machine learning. The final product available thus provides semantic classes such as ground, vegetation, buildings... To render vegetation, only points from the aforementioned class were used. However, an issue quickly arose: topographic features such as roads were misaligned with lidar points cloud, by a 1 or 2 meters shift, creating sometimes an overlap between the two (trees at the middle of the road). This was first observed through trial and error while integrating the two data, and later strengthened by external analysis using QGIS (fig. 2), and then confirmed by discussing with other geospatial experts. This issue was not exclusive to vegetation and roads: It also applied to buildings, defined both by a geometry in the topographic data, and also points classified as building in the lidar point cloud. When using a trivial intersection of the two, it resulted in poor visual quality, because of the misalignment. To tackle this issue, and improve visual representation of buildings, the software roofer[1] was used. It generates LoD 2.2 (Biljecki et al., 2016) buildings by combining roofprints and classified points cloud, with a tolerance to misalignment. An alternative solution could be to alter the buildings geometry to make them match the points from the points cloud, and then use those realigned geometries with a trivial intersection, which should better overlap (article not yet published). This realignment algorithm was not ready to use compared to roofer, so the latter was selected. However, roofer only handles buildings, while this alternative algorithm could be applied to other features, and may be part of the solution regarding roads and vegetation misalignment, which was not solved here. It could also be used as a preliminary step before using roofer, as it may improve the quality of its results.

Problem 4 (documentation accessibility): To properly render buildings, a texture or color is needed alongside the 3D geometry computed by roofer. No terrestrial acquired data were available to provide facades pictures, however the BD TOPO had attributes called ”roof material” and ”wall material”. At first, it seemed possible to use them to choose a texture from a set of predefined ones, to make the buildings look realistic from a distance. The issue was that those attributes were filled only for approximately two thirds of the buildings, leaving too many of them untextured. Instead, a thematic color has been assigned to the roof of each building according to its classification. Such knowledge about the data could have been beneficial to have in first place. Indeed, if the DTT engineer knew that

[1] `https://innovation.3dbag.nl/roofer/`

https://doi.org/10.5194/isprs-annals-XI-1-2026-355-2026 | © Author(s) 2026. CC BY 4.0 License.

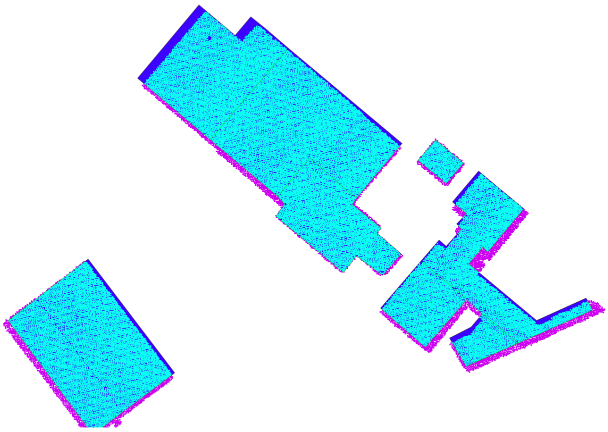

Figure 2. Data misalignment between lidar points (magenta/cyan) and 2D geometries (blue). Overlapping points are shown in cyan, outside in magenta. (Problem 3)

he would not be able to rely on those attributes, he might have chosen another data source more suitable for his goal. But the documentation only states that the attributes are not mandatory, without any further indication on when they are and are not filled, how much is, and what alternative could be used when they are not. Furthermore, the documentation itself is not always easily accessible, as it is a huge PDF file. A useful tool here is BD TOPO Explorer[2] which provides a search box and links to relevant sections of that documentation.

Problem 5 (data quality): One last issue was observed when exploring the result: some buildings were rendered with a clearly incorrect and unrealistic roof geometry. This is the case of the town hall and the church: their feature IDs are respectively: `BATIMENT0000000292502020` and `BATIMENT0000000292501371`. However those errors are not caused by the process used (roofer), but instead originate from the initial raw data (aerial lidar points cloud). The two buildings have a very steep roof part, which has not been properly recorded due to the flight angle, thus making it nearly impossible to reconstruct the 3d representation without any further manual fixing.

A screenshot of the data integration result, taken from its environment, Minetest, can be seen in figure 3.

4.2 Instantiating the model

4.2.1 The recipe The previous section describing the work of the DTT engineer is an instantiation of our model. The first paragraphs describe respectively the territory subject, the DTT engineer agent, and the stake. Those give all contextual information required to understand the situation, before actually going into the technical work of the engineer. The following paragraph (called DTT) is the interpretation, and states the DTT goals that the engineer decided to target.

Paragraphs called Process or Problem represent technical works and issues encountered. Problems are fully described, and some are accompanied by proposed solutions, while others only have hints or even no solution at all. A problem type is also assigned to each of them. Not all types of problems identified were shown here in this recipe.

[2] `https://bdtopoexplorer.ign.fr` (in French)

4.2.2 Discussion This recipe describes a work done one year earlier, and was formally written from notes taken by the DTT engineer at the time of the work. We consider that the time gap does not have a significant impact on the quality of the recipe itself. Nevertheless, we are looking forward to conduct another experiment where the recipe is written concurrently, to evaluate whether it represents a huge workload for DTT engineers. This new experiment would explore another use case, around the renovation of a public library: a building surrounded by a small park.

The workflow is not fully modeled here, with parameters and data links. Thus, this recipe is partial and could not be reapplied as it is. However, it still contains expertise, especially concerning the interpretation of the stakes, and some DTT engineering problems. It could then be completed later, to provide the missing parts, and become a complete recipe that anyone could reuse.

Additionally, in this version, texts are not annotated, making them standard plain texts instead of `richtexts` as described in the section 3.1. A simple editor providing auto-completion of annotations could help to write them. A very early prototype was drafted and looked promising, but needs refinements to better assist the DTT engineer in his writing process.

The making of this recipe involved little to no collaboration. This could be explored in future works. Specifically, we would like to explore how to maintain multiple versions of a recipe, where multiple DTT engineers suggest improvements or new parts to a single recipe.

4.3 Reusing the recipe

One initial goal of the model was to allow sharing DTT engineering work in such a way that other DTT engineers could find and reuse recipes or parts of recipes. To assess this, we need to find volunteers to try to follow a recipe, then compare their results to ours, and collect their feedback on their experience: whether they found the recipe easy to follow or not, what information they were missing to understand a task, or what they would in turn share about this recipe, to further improve it. This experiment could be conducted on multiple recipes, targeting various stakes and levels of expertise. Targeted profiles could include researchers, geospatial experts, or municipal agents with basic geospatial knowledge. It will be part of the continuation of this work.

5. Conclusion

This paper presented a work in progress to integrate engineering practices as a core component of national DTT strategies. Our motivation is to consolidate local experiences into shared and reusable knowledge to improve the alignment of tools and datasets with real user needs, and accelerate the deployment of robust, context-adapted Digital Twins of Territories.

We introduced `dttrecipe`, a model designed to capture and structure the expertise of engineering specific Digital Twins of Territories to meet specific local stakes. The model builds on established provenance and workflow ontologies and drawing on inspiration from the geospatial user-feedback standard to provide a way to document different categories of knowledge: the expression of territorial interests, design choices and the rationale behind them, data selection and transformation, encountered engineering issues and how they were solved.

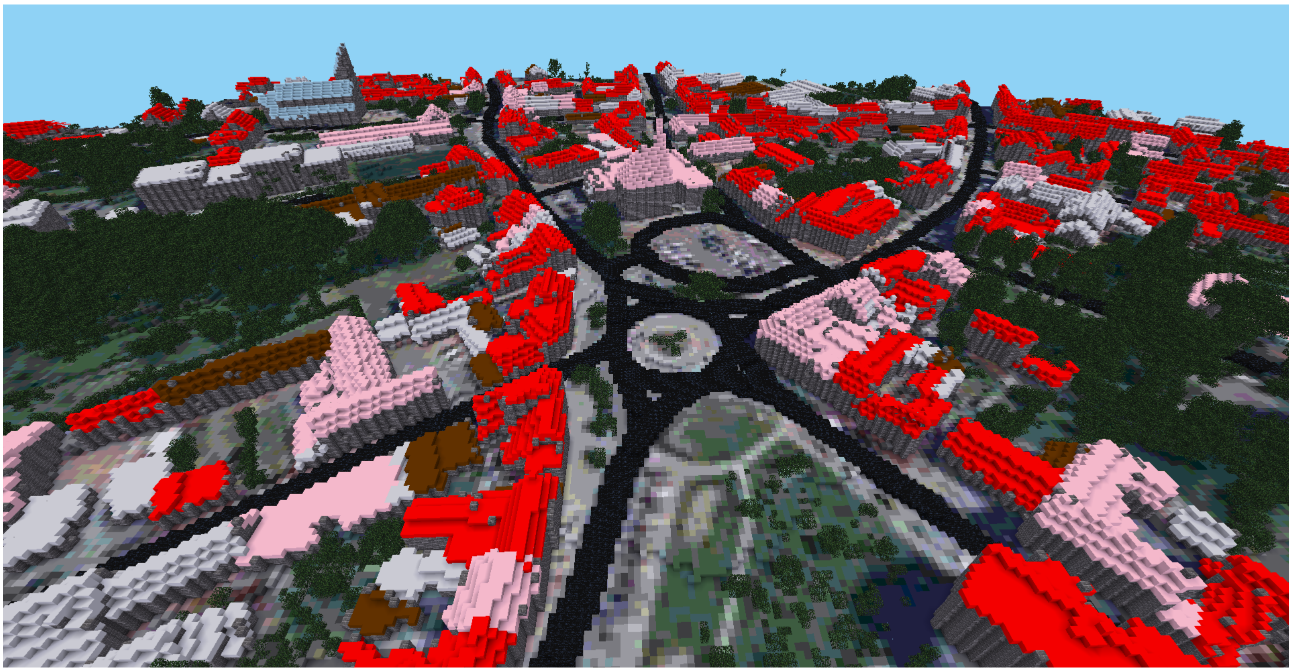

Figure 3. Data integration result, in the Minetest environment

A first case study demonstrated that `dttrecipe` effectively reveals recurrent categories of engineering challenges — such as data discoverability, multi-source inconsistencies, uneven documentation quality, and gaps in local expertise — that strongly influence DTT conception even though they are rarely formalized.

We are preparing the validation of problem types, through a survey targeting all profiles related to DTT data engineering, to gather feedback on the current proposed types and suggestion for new ones. This study will also be a way to accumulate some knowledge about various problems, which will serve as a basis for a first knowledge base to construct recipes from.

Future work will investigate the acquisition of more DTT design processes from technical reports, data papers analysis and tests with the engineers. We believe that DTT engineers are more inclined to share their knowledge about individual problems, rather than full recipes. We should consider ways to aggregate feedback from multiple actors to form a composite recipe. We will also design querying and visualization mechanisms for recipe exploration, across diverse territorial contexts.

We also want to emphasis on the formalization of visualization techniques, especially about 3D and also "invisible" data that are not trivial to integrate (such as statistical data, citizen's feedback, or territorial dynamics). Finally, we would like to expand on the interactivity of DTTs, and how can actors (such as citizens) express themselves through those.

6. Acknowledgments

This work was supported by government funding administered by the French National Research Agency under the France 2030 Plan, reference ANR-21_EXES_007.

References

Agrawal, A., Thiel, R., Jain, P., Singh, V., Fischer, M., 2023. Digital Twin: Where do humans fit in? *Automation in Construction*, 148, 104749.

Béler, C., Zacharewicz, G., Bisgambiglia, P.-A., Poggi, B., Poux, F., Antoine-Santoni, T., 2024. Towards (an Aggregated) Territorial Digital Twin / From smart-village to smart-territory via the territorial system of digital twin. *Concepts in Smart Societies : Next-generation of Human Resources and Technologies*, CRC Press, pp. 328–358.

Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., Mina, E., Corcho, O., Gómez-Pérez, J. M., Bechhofer, S., Klyne, G., Goble, C., 2015. Using a suite of ontologies for preserving workflow-centric research objects. *Journal of Web Semantics*, 32, 16-42.

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, 59, 25-37.

Bleisch, S., 2012. 3D GEOVISUALIZATION – DEFINITION AND STRUCTURES FOR THE ASSESSMENT OF USEFULNESS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-2, 129–134. https://isprs-annals.copernicus.org/articles/I-2/129/2012/.

Bogdanović, M., Stanimirović, A., Stoimenov, L., 2015. Methodology for geospatial data source discovery in ontology-driven geo-information integration architectures. *Journal of Web Semantics*, 32, 1-15.

Breunig, M., Bradley, P. E., Jahn, M., Kuper, P., Mazroob, N., Rösch, N., Al-Doori, M., Stefanakis, E., Jadidi, M., 2020. Geospatial Data Management Research: Progress and Future Directions. *ISPRS International Journal of Geo-Information*, 9(2). https://www.mdpi.com/2220-9964/9/2/95.

Callcut, M., Cerceau Agliozzo, J.-P., Varga, L., McMillan, L., 2021. Digital Twins in Civil Infrastructure Systems. *Sustainability*, 13(20). https://www.mdpi.com/2071-1050/13/20/11549.

Coetzee, S., Ivánová, I., Mitasova, H., Brovelli, M. A., 2020. Open Geospatial Software and Data: A Review of the Current State and A Perspective into the Future. *ISPRS International Journal of Geo-Information*, 9(2). https://www.mdpi.com/2220-9964/9/2/90.

de Andrade, B., Poplin, A., Sousa de Sena, Í., 2020. Minecraft as a Tool for Engaging Children in Urban Planning: A Case Study in Tirol Town, Brazil. *ISPRS International Journal of Geo-Information*, 9(3). https://www.mdpi.com/2220-9964/9/3/170.

Ellul, C., Stoter, J., Bucher, B., Olsson, P.-O., Billen, R., DeLathouwer, B., 2024. Towards National Connected Digital Twins - A Geospatial Perspective. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W5-2024, 147–154. https://isprs-annals.copernicus.org/articles/X-4-W5-2024/147/2024/.

Konstantinidou, M., Grau, J. M. S., Sahala, S., Nesterova, N., de Kruijf, J., 2023. A methodology for using dynamic visualizations to enhance citizens engagement. *Transportation Research Procedia*, 72, 4026-4032. TRA Lisbon 2022 Conference Proceedings Transport Research Arena (TRA Lisbon 2022),14th-17th November 2022, Lisboa, Portugal.

Lecordix, F., Kumarasamy, S., 2021. Teaching on natural risks by simulation in Minetest and augmented reality. *Abstracts of the ICA*, 3, 173. https://ica-abs.copernicus.org/articles/3/173/2021/.

Metcalfe, J., Ellul, C., Morley, J., Stoter, J., 2024. Characterizing the Role of Geospatial Science in Digital Twins. *ISPRS International Journal of Geo-Information*, 13(9). https://www.mdpi.com/2220-9964/13/9/320.

Moreau, L., Groth, P., 2022. *Provenance: an introduction to PROV*. Springer Nature.

Onyimbi, J. R., Koeva, M., Flacke, J., 2018. Public Participation Using 3D Web-Based City Models: Opportunities for E-Participation in Kisumu, Kenya. *ISPRS International Journal of Geo-Information*, 7(12). https://www.mdpi.com/2220-9964/7/12/454.

Scholten, H., 2017. Geocraft as a Means to Support the Development of Smart Cities, Getting the People of the Place Involved - Youth Included -. *Quality Innovation Prosperity*, 21(1), 119–150. https://qip-journal.eu/index.php/QIP/article/view/784.

VanDerHorn, E., Mahadevan, S., 2021. Digital Twin: Generalization, characterization and implementation. *Decision Support Systems*, 145, 113524.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18.

Xavier, T., Derksen, D., Martin, V., Brunet, P.-M., 2025. Building a framework for the design and deployment of digital twins: the Digital Twin Factory project. *EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025*, EGU25-1299. https://meetingorganizer.copernicus.org/EGU25/EGU25-1299.html.

Zabala, A., Masó, J., Bastin, L., Giuliani, G., Pons, X., 2021. Geospatial User Feedback: How to Raise Users’ Voices and Collectively Build Knowledge at the Same Time. *ISPRS International Journal of Geo-Information*, 10(3). https://www.mdpi.com/2220-9964/10/3/141.

Zhang, Y., Pappa, C. I., Pittich, D., 2024. Exploring user-generated content motivations: A systematic review of theoretical perspectives and empirical gaps in online learning. *Computers and Education Open*, 7, 100235.