PED
EU
NET

COST Action 19126

Positive Energy Districts European Network

Deliverable 2.5

Report on Graph-Based Framework for Positive Energy Districts: Integrating KPIs, Tools, Stakeholder Roles, and Intervention Strategies

Authors: Ábel Magyari, Matthias Haase, András Reith

Final delivery date: 15.03.2025

This publication is based upon work from COST Action Positive Energy Districts European Network (PED-EU-NET), supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.cost.eu

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

Executive Summary

This deliverable of Task 2.5 introduces a PED Knowledge Graph that maps

- 242 KPIs (drawn from PED relevant projects in the PED-EU-NET database, namely Syn.ikia, Cultural-E, ARV, SmartEnCity),
- 39 design and simulation tools, six categories of stakeholder roles (spanning legislative, planning/design, implementation, operation, end-use, and finance),
- 18 intervention points across four key intervention domains (Energy, Urban and local development and real estate, Infrastructure and People).

We detail the creation of this knowledge graph from data curation and ontology design to relationship modelling via literature and expert judgement and demonstrate its utility in supporting decision-making. Through queries, we show how the graph enables retrieval of insights such as most impactful tool for a specific stakeholder, or identification of tool gaps in the current PED tool scene.

The results highlight improved interoperability of knowledge across traditionally siloed domains, enabling more informed and collaborative PED design processes. We also evaluate the system's performance (with sub-second query responses on the Neo4j graph) and integrity.

In discussion, we position the PED Knowledge Graph within broader PED design efforts and outline future enhancements, including integration of governance/policy knowledge, inclusion of non-technical "soft" tools, and coupling with Retrieval-Augmented Generation (RAG) frameworks for dynamic tool chain recommendations.

This research contributes a novel, structured approach to capturing and leveraging PED design knowledge, laying groundwork for more transparent, data-driven, and multi-stakeholder decision support in sustainable urban planning.

Table of Content

1. Introduction

Positive Energy Districts (PEDs) are emerging as a cornerstone of sustainable urban development, as urban areas with net positive energy balance and contributions to broader climate and societal goals[1]. By definition, a PED must integrate renewable energy generation to exceed its own consumption, but recent studies emphasize that PEDs also aim to improve quality of life, economic resilience, and social inclusiveness [2]. The PED concept thus extends beyond energy performance alone, requiring a holistic design approach that simultaneously addresses environmental, economic, and social KPIs. Achieving these multifaceted objectives is a complex challenge: PED projects involve multiple stakeholders (e.g. city authorities, planners, developers, citizens), span various scales (building to district), and necessitate interdisciplinary tool workflows for analysis and simulation [3]. Traditional design processes and certification frameworks have struggled to accommodate this complexity. For instance, current neighborhood sustainability certification systems (LEED-ND, BREEAM Communities, etc.) provide valuable criteria and indicators but their diversity in system boundaries as well as metholodogies make comparability and understanding for urban scale developments rather challenging[4].
Recent research by Volpatti et al.[5] underscores the need for a PED-specific evaluation protocol, noting that existing schemes do not fully capture PEDs' potential or unique features. Similarly, Natanian et al. (2024) highlight significant gaps in tool support and integration for PED design, while a variety of simulation and planning tools exist, they often operate in silos, focusing on isolated aspects (energy, climate, etc.) without seamless interoperability or shared knowledge bases[3].

In summary, the state-of-the-art reveals a pressing need for integrative frameworks that connect the diverse knowledge domains of PED design: performance indicators, stakeholder activities, and tool capabilities.

This report aims to address these challenges by developing a PED Knowledge Graph, a structured representation of PED design knowledge implemented on a Neo4j graph database[6]. The knowledge graph approach allows us to map heterogeneous entities, including KPIs, stakeholder roles, design tools, and intervention strategies, and their interrelationships in a single connected data model.
Knowledge graph is defined as a directed graph that encodes real-world facts by representing entities as nodes and their semantic relationships as edges - typically structured in triplets and organized under an ontological schema - to serve as an effective knowledge base [7,8].

By capturing expert knowledge about which tools address which KPIs, which stakeholders are involved in which interventions, etc., the graph provides a unified decision-support resource. Graph databases are well-suited for this task: they represent knowledge as nodes and relationships, enabling intuitive queries across connected data and uncovering multi-hop relationships that would be difficult to trace in conventional databases[9].Furthermore, knowledge graphs have already been successfully applied in related engineering domains to improve decision transparency and multi-objective collaboration [10]. In the context of PEDs, the knowledge graph aims to function as a "single source of truth" linking critical design elements.

We hypothesize that this approach can enhance interoperability (by linking data and tools across domains), support tool selection and chaining (by recommending suitable tools or sequences of tools for given design goals), and facilitate stakeholder communication (by making explicit the connections between stakeholders and important decision making phases, and tools).

The remainder of this report is organized as follows :

In *2.Related work*, we survey existing research on knowledge graphs in engineering and urban sustainability, and we review recent PED literature to establish the context and identify gaps that our work addresses. The *3. Methodology* section then describes the developed framework in detail: the rationale for choosing Knowledged Graphs, the definition of entity types (KPIs, tools, stakeholders, interventions) and their properties, the data collection and curation process, and the modelling of relationships (illustrated with schema diagrams).
In *4. Analysis and Results*, we demonstrate the knowledge graph's capabilities through example queries, showing how insights can be extracted for PED design decision-making. We also present metrics on query performance and discuss validation of the graph's content.
The *5. Discussion* examines the implications of our findings, evaluating how well the current knowledge graph meets PED design needs, and discussing limitations such as scope and data completeness. We then outline future directions, including incorporation of dynamic tool chain recommendation systems leveraging the knowledge graph (possibly via RAG and AI-driven techniques).
Finally, the *6. Conclusion* summarizes our contributions and the significance of a knowledge graph approach for advancing PED design and implementation.

2.Related work

2.1 Knowledge Graphs in Engineering Design and Decision Support

Knowledge graphs (KGs) have gained prominence as a means to integrate heterogeneous data and knowledge in complex domains. A knowledge graph represents information as a network of nodes (entities) and edges (relationships), often enriched with semantic context [11]. Unlike static document repositories or siloed databases, KGs enable flexible querying and reasoning across interconnected facts, making them powerful for decision support in scenarios involving many interdependent factors[10]. Prior research has demonstrated the utility of graph-based knowledge models in engineering and planning disciplines. Mittermeier et al. (2024) proposed a graph-based approach to support group decision-making in manufacturing, using a graph database to capture dependencies between objectives, stakeholders, and solution options in multi-objective optimization problems[10] The knowledge graph in that context improved transparency and traceability of decisions by allowing stakeholders to traverse from high-level goals down to underlying data and models. Similarly, in the building design and operations domain, knowledge graphs and ontologies have been employed to enhance data interoperability, for example, by linking Building Information Modeling (BIM) data with performance simulation results to facilitate cross-tool data exchange [12].

These efforts underscore that KGs can serve as an "interoperability foundation" that bridges multiple tools and data formats, addressing fragmentation in the Architecture, Engineering, Construction (AEC) industry. More broadly, knowledge graphs are seen as a key enabler for integrated analysis in sustainability contexts: Fotopolou et al. (2022) presented *SustainGraph*, a KG for tracking Sustainable Development Goal indicators, arguing that KGs' flexible schema and ability to connect disparate domains make them well-suited for complex socio-environmental systems [9].
This flexibility is also crucial for PEDs, which sits at the nexus of urban and building scale energy systems planning, architectural planning, and optimal indoor and outdoor comfort.

Despite growing interest in knowledge graphs, their application to tool selection and workflow integration in engineering fields remains nascent. Traditional approaches to tool interoperability (such as file exchange standards or direct software integrations) lack a high-level semantic layer to capture *what* each tool contributes to a project and how different tools and outcomes relate.

This work contributes to this area by using a knowledge graph not just as a data integrator, but as a knowledge base for tool capabilities and their linkage to design objectives. By encoding expert knowledge about tools and KPIs, the graph can answer complex questions (e.g., "Which simulation tool can evaluate a given KPI, and which stakeholders are most likely interested in this tool?") that would be cumbersome to resolve without a connected knowledge model.

2.2 Tool Chains and Interoperability in PED Design

Recent literature on Positive Energy Districts has begun to grapple with the challenge of selecting and integrating tools to address the myriad design objectives of PED projects, like deliverable D2.1 from PED-EU-NET.
Natanian et al. (2024) provide a comprehensive overview of "ten questions on tools and methods for PEDs," offering insight into the current landscape and limitations of PED design tools. One key conclusion from the work is that no single tool can cover all PED aspects; instead, an adaptive set of tools is required to tackle diverse domains (energy, carbon, urban climate, mobility, social factors, etc.) [3]. The paper categorizes available PED tools by the domains covered, modeling approaches, and outputs [3].For example, many tools specialize in building- or district-scale energy simulation, while others address environmental externalities or economic analysis. Crucially, Natanian et al. point out that integrated workflows are needed: combining multiple tools in a unified process can yield cross-disciplinary insights that single tools cannot achieve[3]. It mentions the use of platforms like Grasshopper (a parametric design environment) to link tools via plugins, enabling data to flow from one simulation to another. This kind of interoperability is seen as essential to handle PED complexity, allowing, for instance, urban form generation to inform energy simulations, which in turn inform mobility or microclimate models.
However, establishing such interoperability remains challenging; common issues include mismatched data formats[13], inconsistent indicators and need for a standardized way to assess impacts[14], and lack of guidance on which tool should be used at which design stage[3].
This work addresses this gap by providing a higher-level knowledge layer describing relationships among tools and design objectives, which can complement the low-level data interoperability solutions. In essence, while middleware and data schemas handle data

exchange, the knowledge graph handles knowledge exchange, informing designers *what* set of tools might be needed and *how* they relate via shared KPIs or processes.
Another insight from Natanian et al. is the importance of multi-objective and multi-stakeholder considerations in PED tool usage. PED design must incorporate a spectrum of performance metrics aligned with the UN Sustainable Development Goals, and tools need to collectively address this spectrum [3].
Furthermore, engaging various stakeholders through these tools is necessary to ensure that solutions are realistic and accepted. For instance, decision-support tools might be used in participatory workshops to help citizens and local authorities visualize trade-offs. Yet, as highlighted by a PED stakeholder mapping study [15], there is a lack of structured methods to map which stakeholders should be involved at different design phases and how their input connects to technical evaluations.
By including stakeholder roles in our knowledge graph, we attempt to bridge this gap, linking stakeholders to the KPIs they influence and tools they commonly use or interact with.

2.3 PED Indicators and Certification

Defining and tracking KPIs for PEDs is another active area of research. For example to make indicators more inclusive and to uncover trade-offs, Salom et al. introduced an evaluation framework, ensuring multidimensionality during indicator selection [16]. In another publication, Volpatti et al. examined how existing neighborhood sustainability protocols can be adapted into a PED-specific certification methodology . Their analysis of LEED-ND, BREEAM Communities, and CASBEE-UD identified common sustainability criteria, which they then cross-referenced with PED goals to propose additional KPIs relevant to PED certification. For complex systems like PEDs, a more transparent representation of indicators is needed to understand trade-offs. This motivates our approach of treating each KPI as a node in the graph, maintaining the granularity of information. By linking KPIs to interventions and tools explicitly, we can avoid the “black box” of weighted scores and instead trace how a given design strategy might impact multiple KPIs.

In summary, the literature points to significant research gaps that our PED knowledge graph aims to fill.

These include:

- A lack of integrated frameworks connecting PED design knowledge across technical and non-technical domains;
- The difficulty of selecting appropriate tool chains for PED projects in the absence of a structured knowledge base; and
- Insufficient mapping of stakeholder involvement to design interventions and outcomes. While previous works have identified what needs to be done(e.g., achieve interoperability, engage stakeholders, define new indicators) our work operationalizes these concepts into a concrete knowledge graph implementation. By building on insights from the available broader literature, and especially the knowledge gathered in the PED-EU-NET, we ensure that the content and structure of the knowledge graph aligns with current expert understanding of PED challenges, thereby making it a relevant foundation for advanced applications such as decision support systems.

3. Methodology

3.1 Overall Approach and Graph Construction

Our methodological framework centers on constructing a Neo4j[6] property graph that encapsulates key entities and relationships in PED design.

Neo4j was chosen as the platform for several reasons:

- Its labeled property graph model provides an intuitive way to represent domain concepts (each entity can be a labeled node with properties, and relationships can also carry type labels and attributes);
- It offers a powerful query language (Cypher) for traversing complex relationships and pattern matching, enabling the kinds of queries needed for design decision support; and
- Neo4j's ecosystem includes integration with various programming languages and tools, as well as support for graph algorithms, which could be leveraged for future analytics (e.g., centrality to find key connector nodes, community detection among KPIs, etc.).

Another consideration was compatibility with Retrieval-Augmented Generation (RAG) frameworks. In envisioning advanced PED decision-support assistants, we anticipate using the knowledge graph as a retrieval component that provides grounded facts to a language model. RAG is a technique that combines information retrieval with large language model generation to improve the factual accuracy of AI responses[11].
By structuring PED knowledge in Neo4j, a queryable knowledge base was created that an LLM can draw from, thereby reducing hallucination and injecting domain-specific accuracy into the generative process [11].

Entity Definitions and Data Sources

We defined four primary entity types (node labels) in the PED knowledge graph: KPIs, Tools, Stakeholder roles, and Intervention points. Each corresponds to a fundamental concept in PED planning:

- **KPIs:**

  This node type represents a performance metric or criterion used to evaluate PED outcomes. We populated 242 KPI nodes, drawing from multiple European project sources that have defined PED or sustainable neighborhood KPIs.
  Key sources included the Syn.ikia project [2] (which focuses on Sustainable Plus Energy Neighborhoods), Cultural-E[17], and ARV[18]among others. These projects provided a rich list of indicators across categories such as energy (e.g. annual net energy balance, renewable ratio), environmental (e.g. CO_2 emissions, air quality index), economic (e.g. life-cycle cost, return on investment), social (e.g. affordability index, user satisfaction) and others. We ensured to include KPIs related to energy flexibility (e.g. peak load reduction, demand response potential) and inclusiveness/quality of life (e.g. fuel poverty rate, accessibility, urban greening index)

because literature highlights these as emerging PED considerations [14]. For now, KPIs only have a name property.

- **Tools:**

  This node type encapsulates software tools, simulation engines, or methods used in the PED design process. We included 39 Tool nodes, collected from Cost Action deliverable D2.1 and Natanian et.al.[3]. Examples of tools in our graph include energy simulation software (e.g. EnergyPlus, TRNSYS, City Energy Analyst), optimization frameworks (e.g. EnergyPLAN for district energy scenarios), urban climate modeling tools (e.g. ENVI-met for microclimate), mobility and transportation models, and decision-support tools (e.g. multi-criteria analysis software.

- **Stakeholder_roles:**

  This node type represents generic stakeholder categories involved in PED design and implementation. Rather than listing individual stakeholders or organizations, we focused on role categories derived from Li et al based on urban development stakeholder analysis[19]. We identified six broad role categories that map to phases of the project lifecycle and interest areas:

  1. **Legislative** e.g. local authorities, urban planners in municipal government, policy makers who set enabling frameworks.
  2. **Planning and Design:** e.g. architects, urban designers, engineers, consultants responsible for the PED concept and technical design.
  3. **Construction and Installation:** e.g. construction firms, technology providers, contractors who realize the PED infrastructure.
  4. **Operation and Maintenance:** e.g. facility managers, energy service companies, utilities handling the operation of energy systems and upkeep of assets.
  5. **Connected to energy end use :** e.g. residents, tenants, local community, and potential energy community members who use or provide energy and are directly impacted by PED performance.
  6. **Stakeholders connected to financial aspects:** e.g. funding agencies, investors, project developers interested in the economic returns and financing of PED projects.

- **Intervention points:**
  This node type corresponds to major decision points in PED design [3]. Essentially these are the thematic aspects in which multiple stakeholders have to navigate and make decisions in. Based on the relevant aspects collected in Volpatti et. al[5] 18 distinct intervention points were identified. These intervention points are categorized into four categories: Energy, Urban and local development and real estate, Infrastructure, People.
  These categories were chosen to cover both the technical and non-technical aspects of PED interventions.

Relationship Modeling

With the nodes defined, we established relationships between them to encode expert-judgment links. The relationship types in our knowledge graph include:

- **be_calculated_by** (KPI → Tools): Perhaps the most critical relationship type, this link indicates that a given KPI can be evaluated or calculated by a certain Tool. For example, a KPI like „*Non-renewable primary energy consumption*" would have a *be_calculated_by* link to building energy simulation tools, and KPIs like *Modal Share of Sustainable Transport* might be adressed by urban mobility tools. We populated these links through mapping KPIs to tools based on tool documentations (what outputs or analysis each tool provides).
- **uses** (Stakeholder_roles → Tools): This relationship denotes which stakeholder roles typically use or operate a given tool. For instance, a Planning and Design professional might *uses* energy simulation software, whereas a policy planning tool could be *USED_BY* "Legislative" stakeholders. Some tools can have multiple users; e.g., a visualization dashboard might be used by both technical experts and presented to end-users. This link helps identify capacity-building needs and collaboration points – if a tool needed for a certain analysis is not in the typical toolkit of a certain stakeholder, that gap becomes evident.
- **connected_to** (KPI → Stakeholder_roles): We introduce this relationship to capture which KPIs are of particular interest or concern to each stakeholder category. For example, the Legislative role will be *connected_to* KPIs like "Affordability of housing" or "Legal framework compatibility", while the *Operation and maintenance* role will be *connected_to* felxibility KPIs (*Net energy duration curve, Electricity demand load profile*). This layer of the graph connects the human factor to performance metrics, allowing queries that trace, for instance, "*which stakeholders care about the KPI X?*" or "*which KPIs does stakeholder Y influence?*".
- **belongs_to** (KPI → Intervention points): This relationship maps interventions points to the KPIs they primarily contain. It reflects design intent: for instance, KPIs such as *Grid purchase factor,* and *Peak power* will *belong_to* the *Energy Flexibility* intervention point, while others like *ICT response time*, and *Improved interoperability* will *belong_to* the *Technological solutions* intervention point. As many KPIs are connected to multiple intervention points, there are some overlaps in the graph, which in turn is creating the connection between intervention points. This part of the graph effectively encodes a conceptual theory of change: enhancing the KPIs associated with specific intervention points lead to improvements in those aspects of the PED.
- **uses_tool** (Intervention points → Tool): This link connects intervention points with tools that can be used to plan, or resolve them. For example, the *Quality of life* intervention point is connected to mainly IEQ indicators, thus it might *uses_tool* that is able to calculate those KPIs like Butterfly, Urbano or Intelligent Community Design (ICD); to evaluate Energy efficiency intervention point, stakeholders might select

urban or building level energy simulation softwares (i.e. EnergyPlus, ESP-r, UMI, Pollination). This relationship shows the operational side – given a strategy domain, what tools help realize it?

- **involved_in** (Stakeholder_roles → Intervention points): This denotes which stakeholder roles need to be involved for a given intervention point. For instance, *Legislative* roles and roles *Connected to energy end use* are critical for intervention points like *Resilience and security of energy supply*, whereas *Operation and maintenance, Planning and design, Construction and Installation* roles and those *Connected to energy end use* must be involved in decisions around *Energy production* intervention point. This relationship, along with the two above, forms a connected loop linking stakeholders, interventions, tools, and KPIs.

Figure 1*. illustrates the schema* of the PED knowledge graph, showing the node types and relationship types described above. It has to be noted, that for operational purposes KPIs are separated into two groups, those that can be calculated with the current tools without any further operation needed from the user side, and those that cannot be calculated or can only be calculated with further operation from the user side.

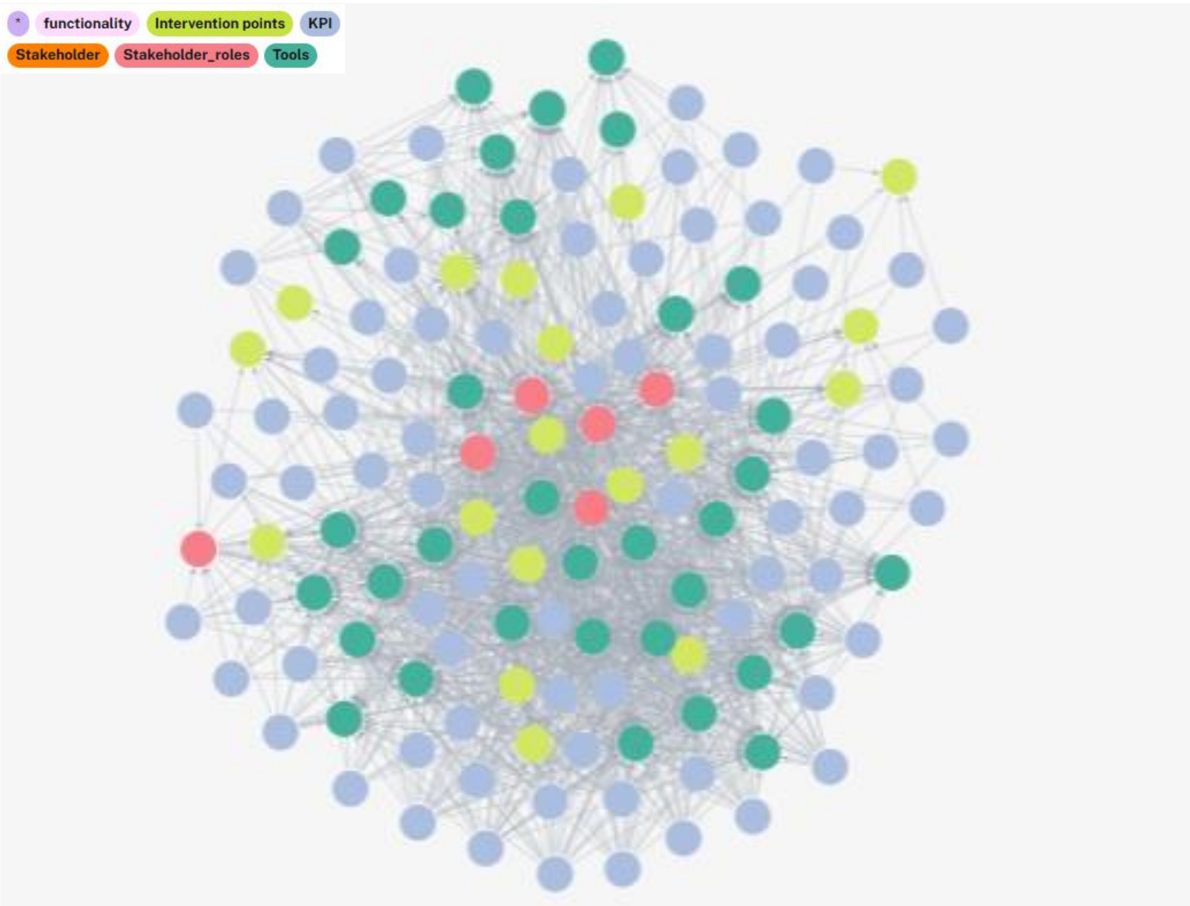

Figure 1. Overview of the PED Knowledge Graph, depicting node types (KPI, Tools, Stakeholder_roles, Intervention points) and key relationships among them

The process of relationship instantiation is based on the available literature and expert judgement. We started by cross-tabulating tools and KPIs based on tool documentation and use cases from PED projects. After, KPIs were also connected to Intervention points based on available categorisation in Volpatti et al[5]. Lastly, Intervention points were connected with relevant stakeholders based on literature[3,19] and expert opinion.

This knowledge curation, utilizing both literature and expert insights, ensured that the graph is neither a purely data-driven construction nor an unbounded brainstorming, but a focused representation of consensus knowledge up to date with PED research.

Finally, we note that the developed graph model is designed to be easily extended. Neo4j's schema-less nature means new node types or relationships can and should be added. This flexibility is important because, PED as a concept is evolving; the knowledge graph is designed to evolve with it by accommodating new knowledge rather than needing a redesign.

4. Analysis and Results

In the following chapter the KGs capabilities to generate insights through queries are showcased to illustrate it's role in driving and supporting decision making. Performance metrics of the graph database and the results of our validation exercises are also summarized.

4.1 Knowledge Graph Querying

The strength of representing PED design knowledge as a graph lies in the ability to ask complex questions that span multiple entities. Neo4j's native Cypher query language is used to interrogate the database. Below are a few examples of queries and the corresponding findings:

- *Query 1: Find the most impactful Tools for a specific Stakeholder.* – This basic query helps identify the tools that have the ability to handle most of the KPIs relevant for a specific stakeholder. For example, if the Specific stakeholder is a *Legislative* stakeholder, the query

  ```
  MATCH (s:Stakeholder_roles {Node:”Legislative”})<-[:connected_to]- (k:KPI)-
  [:be_calculated_by]-> (t:Tools)
  RETURN t.Node, AS Tools, count(k) AS ImpactScore ORDER BY ImpactScore DESC
  ```

  returns a list of tools in descending order showcasing the importance of those for *Legislative* stakeholders, such as *"Intelligent Community Design (ICD), DesignBuilder, CityBES"* indicating that these tools are capable of analyzing or contributing to the most relevant KPIs for Legislative stakeholders. Such a result directly informs the stakeholders about which tools to consider when aiming to evaluate and improve certain KPIs.

- *Query 2: Find the most impactful PED aspect/intervention point.* – This query will help find the most important aspects/intervention points for the PED design, or selected KPIs. For instance, we run

  ```
  MATCH (k:KPI)-[:belongs_to]->(ip:`Intervention points`) RETURN ip.Node AS
  InterventionPoint, count(k) AS ImpactScore ORDER BY ImpactScore DESC
  ```

to get a mapping of the most impactful Intervenion points or aspects for PED design. The result show that *Energy efficiency* and *Quality of Life* are the most important design aspects, followed by *Energy production* and *Energy flexibility*. This kind of output helps identifying which aspects need to most attention, and helps in allocating resources in order the reach the best PED outcome possible.

- *Query 3: Identify intervention strategies for a given stakeholder concern.* – Consider an Urban Planner (Planning and Design) concerned with reducing *Total GHG emissions*. We ask: what intervention points should they pay most attention to and which tools should they consider? A query like

  ```
  MATCH (s:Stakeholder_roles {Node:"Planning and Design"})<-[:connected_to]-(k:KPI
  {Node:"Total GHG emissions"}) MATCH (k)-[:belongs_to]->(ip:`Intervention points`)
  OPTIONAL MATCH (ip)-[:uses_tool]->(t:Tools) RETURN ip.Node AS InterventionPoint,
  collect(DISTINCT t.Node) AS Tools
  ```

  yields, for example, that the *"Nearly zero energy buildings and net-zero energy districts"* intervention point is the most relevant, with tools such as urban simulation tools (e.g. City Energy Analyst, URBANopt, SimStadt, and Urban Modeling Interface), building simulation tools (e.g. EnergyPlus, eQUEST, DesignBuilder, and DOE-2), and optimization/energy system modelling tools (e.g. DER-CAM, Calliope, oemof, and energyPRO) being suggested. This answers in one query what might otherwise require searching through multiple documents – it provides a quick mapping from a stakeholder's high-level goal to concrete interventions and tools.

- Query 4: *Finding tool gaps for further development - To look for tool gaps, one could query the intervention points with the most KPIs that are not calculable with the currently collected tools. A query here could be :*

  ```
  MATCH (ip:`Intervention points`)<-[:belongs_to]-(n:NonCalculableKPI)
  WITH ip, count(n) AS nonCalcCount, collect(n.Node) AS nonCalcKPIs
  ORDER BY nonCalcCount DESC
  LIMIT 3
  RETURN ip.Node AS InterventionPoint, nonCalcCount AS NonCalculableKPICount,
  nonCalcKPIs AS NonCalculableKPIs
  ```

  This query for example yields the 3 intervention points with the most KPIs that are not covered by the current list of tools. These are the *'Quality of life'* with 23 such KPIs, the *' Green and blue infrastructures'* with 19 KPIs, and lastly the *'Energy flexibility'* with 11 connected KPIs. This answer might help tool developers in driving their attention to developing tools and solutions that focus on these areas, rather than further improve and develop tools in well-represented areas.

The above queries are just illustrative; many more complex questions can be answered (e.g., "Which stakeholder roles are the most central in the graph?" using graph algorithms to see who is most connected).

4.2 Performance and Validation

The PED knowledge graph, in its current form, contains on the order of a few hundred nodes and a few thousand relationships (each KPI connected to multiple tools and interventions, etc.). This size is relatively modest for Neo4j, and our tests indicate that query performance is excellent – simple queries return almost instantaneously (<50 ms), and even more complex pattern queries with multiple hops execute in under a few hundred milliseconds. This responsiveness is important for interactive use and also bodes well for integrating with real-time applications (where the graph can be queried live without noticeable delay to the end-user).
Validation of the graph's content was conducted by performing consistency checks using Cypher queries (e.g., ensuring every KPI is linked to at least one Intervention, so that no KPI is orphaned without an addressing mechanism). These checks passed, confirming that our curated relationships covered all items (for instance, if a KPI had initially no tool, we revisited sources to find if perhaps a tool exists, or flagged it as a gap for future research).

In summary, the analysis demonstrates that the PED knowledge graph not only accurately encapsulates known relationships but is also a practical instrument for gleaning insights. By answering complex questions, it provides value beyond a static list of KPIs or tools. These are essential capabilities for moving PED concepts from theory to implementation on the ground.

5. Discussion

The development of the PED knowledge graph reveals both the promise of knowledge-driven approaches to sustainable urban design and the challenges that come with capturing a complex domain. In this section, the current model's strengths and limitations is assessed and how it addresses the gaps in PED planning is detailed. We also consider how this work can be extended in the future to further support PED implementation and similar interdisciplinary design problems.

5.2 Strengths and Contributions

A primary strength of the PED knowledge graph is its ability to integrate diverse domains of knowledge into a single, queryable framework. Our results show that information traditionally scattered across reports (KPIs in one document, tools in another, stakeholder analysis in yet another) can be linked and accessed in a coherent way. This integrative capability directly addresses the interoperability challenge highlighted by prior researchers[3] not just at a data exchange level, but at a knowledge level. By formalizing the relationships between tools and KPIs, our work provides a structured framework that reorganizes established concepts into an integrated model for PED design. Rather than creating entirely new terminology, it leverages the existing language of policy, stakeholder roles, and performance indicators—clearly mapping how a policy can influence stakeholders, who then engage with specific tools that ultimately affect various KPIs. This unified representation serves as a lightweight decision-support tool, enabling tool developers, urban planners, and policy makers to understand and communicate the cascading effects of changes within the PED system.
In practical terms, the knowledge graph improves traceability. One can trace for example how a high-level goal (like energy flexibility) breaks down into specific indicators and then to tools, which is important for transparent decision-making. This traceability meets the need

for clarity in certification or evaluation protocols and can strengthen the ability to clearly link PED aims and results with higher level national climate goals and sustainable development goals, which is mentioned as a gap in Volpe et al. [4].

Another important contribution is the graph's role in tool and later on tool-chain recommendation. Through simple queries (KPIs ->Tools) graph can suggest tools tool for specific indicators. The graph essentially serves as a knowledge-based recommender system, which could be further developed into a user-friendly application (for instance, an interface where users tick their target KPIs and stakeholders, and the system outputs recommended tools, processes, and stakeholders).

Because of the graph format, graph algorithms or even semantic reasoning could be applied (if extended with ontological semantics). For example, one could run inference rules: if an intervention point has a *belongs_to* relationship with a KPI that is of interest of a certain stakeholder, one might infer that stakeholder should be involved in that intervention point. Some of these inferences we encoded manually as relationships, but in future a reasoning engine could maintain consistency as the graph grows. This points to a direction where the KG could become not just a static repository but an active knowledge base that can answer new questions or detect inconsistencies as the PED domain evolves.

5.2 Limitations

Despite its utility, the current PED knowledge graph has several limitations. Firstly, the relationships are based on *expert judgment and literature as of now*, which introduces subjectivity. There is a risk of biases – for instance, tools that are well-known to authors might be over-represented in links, while less familiar (but potentially useful) tools are under-represented.

Similarly, stakeholder interests and goals were generalized based on literature; in reality, priorities differ by local context (e.g., one community might value aesthetics as part of quality of life, which we did not explicitly include). Thus, the graph, while comprehensive, is not exhaustive or globally objective. It should be seen as a living hypothesis of how things connect, to be tested and refined during usage, especially based on real projects.

Another limitation is the static nature of the data in its current form. The graph captures relationships largely as binary (exists/doesn't exist). It doesn't capture dynamic aspects such as temporal changes or quantitative magnitudes. For example, it can't tell directly how much a given tool will improve a KPI (which would require data, not just a relationship). Nor does it update as a project progresses (unless manually maintained, the graph won't reflect which interventions have been implemented and what outcomes were measured). Incorporating dynamic data (e.g., from monitoring a PED) into the graph is a future step, effectively moving towards a type of digital twin concept [20] where live data feeds the knowledge base. At present, it is a planning tool rather than a real-time monitoring tool.

Additionally, currently "non-technical tools" are omitted. The graph currently shines in linking simulation/analysis tools to KPIs; it is not detailed in linking, say, policy instruments or governance tools to outcomes.

From a user perspective, another consideration is usability: querying a graph database requires some expertise. While the back-end is powerful, an end-user might need a more guided interface. We addressed this partially by creating an MVP of a user-friendly application helping in tool selection based on selected KPIs (However as current graph doesn't allow for evaluation of tool coupling and tool chaining, its usability is limited).

Finally, on the limitation of the validation performed: structure and plausibility was validated, but the ultimate test is usage in real projects. The true measure of success would be to see if teams that use the graph make better decisions (however defined) or avoid certain pitfalls compared to teams that don't. That evaluation is future work; at this stage, only a qualitative validation is provided. Thus, one limitation is that the benefits we claim (improved decision-making, interdisciplinary collaboration) are supported by logical reasoning, but not yet by empirical evidence from project outcomes.

5.3 Comparison to Related Approaches

It is worth contrasting our knowledge graph approach with alternative or traditional methods:

- *Versus Document-based Knowledge Repositories:* A traditional way to gather this knowledge might be a handbook or database that lists KPIs, tools, etc. The knowledge graph has an advantage in that relationships are first-class citizens; you can query across them easily. However, a handbook might provide deeper narrative context for each relationship that our graph doesn't explicitly encode. In practice, one might use the graph in conjunction with narrative guidelines, e.g., once a relationship is identified via the graph, refer to guidelines or case studies for how to implement it.

- *Versus AI without structured knowledge:* With the rise of LLMs, one might ask if an AI chatbot could answer the same questions without a knowledge graph. LLMs can ingest text and give answers, but as noted earlier, they may hallucinate or miss domain nuances [11]. Our RAG-oriented approach of using the knowledge graph aims to combine the best of both: the breadth of AI with the precision of a structured KG. On its own, the KG ensures that any answer is grounded in an actual link or data we've encoded, thus improving reliability.

5.4 Implications for PED Design Practice

If adopted, the PED knowledge graph could influence practice in several ways. It promotes systems thinking[21]– encouraging designers to think about how building-level decisions connect to district-level outcomes and stakeholder values. It could also serve as an educational tool: newcomers to PED projects can navigate the graph to learn the landscape (like a map of what matters in PEDs and how things interrelate). Over time, as more projects use and contribute to the graph, it could become a community resource, capturing best practices (for example, if a new tool is found effective for something, it gets added, slowly shifting the collective understanding). This aligns with the notion of a "collective intelligence" for PEDs[22].

However, caution has to be drawn, that a knowledge graph is not a substitute for human judgment or creativity in the design process, rather it is a support tool. Design problems like PEDs are ill-defined and context-rich; the KG can provide options and highlight considerations, but it doesn't automatically give the "right" answer (indeed, in many cases there isn't a single right answer). Thus, practitioners should use it as a guide and inspiration, but still conduct detailed contextual analysis and stakeholder engagement to make final decisions.

5.5 Future Work

The current PED knowledge graph lays a foundation, but there are numerous opportunities to expand and deepen its capabilities in future research. Below a few directions are outlined:

Future research will utilize the graph's ability to perform multi-step queries to generate dynamic, context-specific recommendations for tool chains. By incorporating detailed information on how tools work together—and integrating AI methods like case-based reasoning—the system could automatically recommend the best sequences of tools to calculate KPIs, considering both their interoperability and calculation features. This flexible approach would customize workflows to the specific needs of each PED project, simplifying decision-making and boosting overall efficiency.

Another possibility is to evolve the current Neo4j property graph into a semantically enriched model by integrating widely recognized ontologies. By mapping KPIs, tools, and stakeholder roles to standard ontological frameworks, the graph would achieve improved interoperability and internal consistency. Enhanced semantic reasoning would empower the system to automatically infer implicit relationships between nodes. For instance, if several tools demonstrate overlapping impacts on key performance indicators, the ontology could automatically suggest them as complementary options during the design process. Furthermore, combining ontology alignment with a reasoning engine would enable the system to deliver recommendations that take regulatory and contextual nuances into account, thereby ensuring both robustness and adaptability. This level of semantic integration would ultimately facilitate the development of a truly dynamic certification framework, one that evolves in step with changing PED requirements and technological advances.

Interoperability remains a critical challenge in the design of PEDs, largely due to the disparate data formats and exchange protocols used across various tools [23] in energy simulation, urban planning, and optimalization. To address with enhancements, the proposed knowledge graph could in the future be transformed into an omni-directional data ontology translation engine. This enhanced system leverages advanced semantic reasoning and tool use[24] to map, translate, and harmonize the diverse ontologies employed by different tools. Specifically, by defining clear input and output standards for each tool within the knowledge graph, unified data model that acts as a common language for all key metrics and data exchanges could be developed. This model would enable seamless omni-directional translation between differing data representations. For example, if one simulation tool outputs data according to ontology A and another planning tool requires input formatted under ontology B, the engine would automatically translate between the two, ensuring that both tools "speak" the same language. Such an approach not only minimizes manual data processing and reduces errors arising from incompatible formats but also underpins dynamic certification processes. By continuously aggregating and analyzing real-time data from multiple sources, the system can offer adaptive, context-aware recommendations that factor in regulatory and situational nuances. In essence, this omni-directional data ontology translation tool would serve as the central engine for data model harmonization, driving both improved tool chain recommendations and a robust, evolving certification framework that keeps pace with advancing PED requirements and technological innovations.

The final, yet equally critical, aspect of future work is real-world validation. Deploying the enhanced knowledge graph in pilot projects or living labs will be essential to assess its practical utility. By embedding the system into actual PED workflows, we can collect feedback on its performance in real-world settings. Metrics such as the reduction in planning time, improvements in KPI performance, and enhanced stakeholder collaboration could serve as indicators of success. Real-world testing would also allow for the iterative refinement of both the dynamic recommendation engine and the semantic reasoning components. As practitioners use the system, their experiences can be used to fine-tune the underlying algorithms and ontology mappings. This continuous improvement cycle will ensure that the knowledge graph remains responsive to emerging challenges and new technological developments in PED design.

In summary, the future work outlined here aims to transform the current PED knowledge graph into a robust, dynamic decision-support system. By focusing on dynamic tool chain recommendations, enhanced semantic reasoning, and interoperability solutions, validated through real-world deployment, we envision a platform that not only supports but also actively evolves with the needs of sustainable urban planning.

6. Conclusion

This document has presented a comprehensive framework for a Positive Energy District knowledge graph, implemented in Neo4j, that structurally ties together the critical components of PED design : KPIs, tools, stakeholder roles, and intervention strategies. Literature collected in the PED-EU-NET augmented with available other relevant literature has been screened to ground our work in the identified needs: bridging interdisciplinary gaps in PED projects, improving tool interoperability, and encompassing the full spectrum of PED goals (energy, environment, social, economic). The resulting knowledge graph encapsulates expert-curated relationships among 242 performance indicators, 39 tools, 18 intervention points, and six stakeholder categories, making explicit the often tacit links that drive decision-making in complex urban sustainability initiatives.

Our analysis demonstrates that the knowledge graph can successfully answer multifaceted queries that would be hard to resolve via traditional means, thus validating the approach as a powerful decision-support tool. By querying the graph, a PED stakeholder can swiftly retrieve information such as which tool to use for a desired outcome, which stakeholders to involve for a given decision, and how various design actions interrelate. Furthermore, the future integration of this graph with a Retrieval-Augmented Generation paradigm suggests a path towards intelligent assistants for PED planning, ones that can provide grounded, context-specific recommendations by leveraging the structured knowledge compiled in this work.

The PED knowledge graph addresses several research gaps: it provides an interoperability scaffold that links previously siloed knowledge domains; it offers a formal yet flexible representation that can evolve with the PED field; and it operationalizes broad concepts (like "holistic design" and "collaborative planning") into tangible connections usable by practitioners. By doing so, it contributes to both theory and practice : theoretically, it exemplifies how knowledge graphs can enhance systems thinking in urban engineering, and practically, it delivers a tool that can be immediately experimented with in PED projects.

However, it has to be also acknowledged that this work is a step in an ongoing journey. PEDs, and sustainable cities more broadly, are dynamic ideas with emerging challenges. A knowledge graph must be maintained as a living knowledge base to remain relevant. Our proposed future directions, including incorporating governance aspects and learning from project deployments, will be crucial for keeping the knowledge graph in sync with real-world needs. Additionally, broadening the scope to include the softer dimensions of planning (like community engagement processes) and linking to real-time data will enrich the utility of the graph.

In conclusion, the Neo4j-based PED knowledge graph represents a novel and timely contribution to the domain of sustainable urban design. It exemplifies how advanced data management techniques (knowledge graphs) can be harnessed to tackle the complexity of PEDs, turning a multitude of data points and expert insights into a cohesive decision support system. As cities and researchers strive to meet ambitious climate and energy targets, tools that improve understanding, communication, and integration will be increasingly important[3]. We envision that the approach detailed in this paper, and its future iterations, will support the creation of Positive Energy Districts that are not only energy-positive, but also well-planned, inclusive, and resilient, ultimately contributing to the broader goal of sustainable, livable cities for all.

References

1. Sareen, S.; Albert-Seifried, V.; Aelenei, L.; Reda, F.; Etminan, G.; Andreucci, M.-B.; Kuzmic, M.; Maas, N.; Seco, O.; Civiero, P.; et al. Ten Questions Concerning Positive Energy Districts. *Build Environ* **2022**, *216*, 109017, doi:https://doi.org/10.1016/j.buildenv.2022.109017.
2. Salom, J.; Tamm, M.; Pascual, J.; Civiero, P. *Syn.Ikia - WP3 Technology Integration in Smart Managed Plus Energy Buildings and Neighbourhoods. D3.1 METHODOLOGY FRAMEWORK FOR PLUS ENERGY BUILDINGS AND NEIGHBOURHOODS*; 2020;
3. Natanian, J.; Guarino, F.; Manapragada, N.; Magyari, A.; Naboni, E.; De Luca, F.; Cellura, S.; Brunetti, A.; Reith, A. Ten Questions on Tools and Methods for Positive Energy Districts. *Build Environ* **2024**, *255*, 111429, doi:10.1016/j.buildenv.2024.111429.
4. Volpe, R.; Bisello, A.; Tuerk, A.; Guarino, F.; Giancola, E.; Sanchez, M.N.; Tumminia, G.; Marrasso, E.; Pallotta, G.; Cutore, E.; et al. Linking Environmental Impact Assessment and Positive Energy Districts: A Literature Review. *Cleaner Environmental Systems* **2025**, *16*, 100264, doi:10.1016/j.cesys.2025.100264.
5. Volpatti, M.; Mazzola, E.; Bottero, M.C.; Bisello, A. Toward a Certification Protocol for Positive Energy Districts (PED). A Methodological Proposal. *TeMA Journal of Land Use, Mobility and Environment* **2024**, *2024*, 137–153, doi:10.6093/1970-9870/10301.
6. Neo4j Documentation - Neo4j Documentation - Accessed 03-16-2025 2025.

7. Bordes, A.; Weston, J.; Collobert, R.; Bengio, Y. Learning Structured Embeddings of Knowledge Bases. *Proceedings of the 25th AAAI Conference on Artificial Intelligence, AAAI 2011* **2011**, 301–306, doi:10.1609/aaai.v25i1.7917.
8. Ehrlinger, L.; Wöß, W. Towards a Definition of Knowledge Graphs. *CEUR Workshop Proc* **2016**, *1695*, 1–5.
9. Fotopoulou, E.; Mandilara, I.; Zafeiropoulos, A.; Laspidou, C.; Adamos, G.; Koundouri, P.; Papavassiliou, S. SustainGraph: A Knowledge Graph for Tracking the Progress and the Interlinking among the Sustainable Development Goals' Targets. *Front Environ Sci* **2022**, *10*, 1–20, doi:10.3389/fenvs.2022.1003599.
10. Mittermeier, L.; Senington, R.; Bandaru, S.; Ng, A. Knowledge Graphs for Supporting Group Decision Making in Manufacturing Industries. *Advances in Transdisciplinary Engineering* **2024**, *52*, 464–474, doi:10.3233/ATDE240189.
11. Vo, N. Mapping the Mind : Knowledge-Graph Augmented Retrieval. **2024**.
12. Wang, Z.; Ying, H.; Sacks, R.; Borrmann, A. CBIM: A Graph-Based Approach to Enhance Interoperability Using Semantic Enrichment. **2023**, 1–10.
13. Malhotra, A.; Bischof, J.; Nichersu, A.; Häfele, K.H.; Exenberger, J.; Sood, D.; Allan, J.; Frisch, J.; van Treeck, C.; O'Donnell, J.; et al. Information Modelling for Urban Building Energy Simulation—A Taxonomic Review. *Build Environ* **2022**, *208*, 108552, doi:10.1016/j.buildenv.2021.108552.
14. Kozlowska, A.; Guarino, F.; Volpe, R.; Bisello, A.; Gabaldòn, A.; Rezaei, A.; Albert-Seifried, V.; Alpagut, B.; Vandevyvere, H.; Reda, F.; et al. Positive Energy Districts: Fundamentals, Assessment Methodologies, Modeling and Research Gaps. *Energies (Basel)* **2024**, *17*, doi:10.3390/en17174425.
15. Cheng, C.; Aelenei, L.; Vandevyvere, H.; Sanchez, M.N. A Systematic Approach towards Mapping Stakeholders in Different Phases of PED Development - Extending the PED Toolbox A Systematic Approach towards Mapping Stakeholders in Different Phases of PED Development – Extending the PED Toolbox. **2021**.
16. Salom, J.; Tamm, M.; Andresen, I.; Cali, D.; Magyari, Á.; Bukovszki, V.; Balázs, R.; Dorizas, P.V.; Toth, Z.; Mafé, C.; et al. An Evaluation Framework for Sustainable plus Energy Neighbourhoods: Moving beyond the Traditional Building Energy Assessment. *Energies (Basel)* **2021**, *14*, 4314, doi:10.3390/en14144314.
17. Al Waheed Hawila, A.; Pozza, C. Cultural-E Project Deliverable D4.1 Evaluation Framework for PEBs. **2022**.
18. Salom, J.; Maskova, I.; Grazieschi, G.; Woods, R.; Schneider-marin, P.; Vaz, D.; Dijkhuizen, M. Van; Andresen, I. ARV-D2 .1 Assessment Framework for Cpcc Wp 2 Framework and Tools for Effective Implementation. **2022**, 1–90.
19. Li, Y.; O'Donnell, J.; García-Castro, R.; Vega-Sánchez, S. Identifying Stakeholders and Key Performance Indicators for District and Building Energy Performance Analysis. *Energy Build* **2017**, *155*, 1–15, doi:10.1016/j.enbuild.2017.09.003.
20. Caprari, G.; Castelli, G.; Montuori, M.; Camardelli, M.; Malvezzi, R. Digital Twin for Urban Planning in the Green Deal Era: A State of the Art and Future Perspectives. *Sustainability (Switzerland)* **2022**, *14*, doi:10.3390/su14106263.
21. Castro, C. Systems-Thinking for Environmental Policy Coherence: Stakeholder Knowledge, Fuzzy Logic, and Causal Reasoning. *Environ Sci Policy* **2022**, *136*, 413–427, doi:10.1016/j.envsci.2022.07.001.
22. Zafeiropoulos, A.; Fotopoulou, E.; Papavassiliou, S. Participatory Socio-Environmental Systems Modeling over Knowledge Graphs. *2021 IEEE Globecom Workshops, GC Wkshps 2021 - Proceedings* **2021**, doi:10.1109/GCWkshps52748.2021.9682047.

23. Dabirian, S.; Saad, M.M.; Hussain, S.; Peyman, S.; Rahimi, N.; Monsalvete Alvarez U, P.; Yefi, P.; Eicker, U. Structuring Heterogeneous Urban Data: A Framework to Develop the Data Model for Energy Simulation of Cities. *Energy Build* **2023**, *296*, 113376, doi:10.1016/j.enbuild.2023.113376.
24. Ruan, J.; Chen, Y.; Zhang, B.; Xu, Z.; Bao, T.; Du, G.; Shi, S.; Mao, H.; Li, Z.; Zeng, X.; et al. TPTU: Large Language Model-Based AI Agents for Task Planning and Tool Usage. **2023**.