

Cross-Domain Flood Risk Assessment for Smart Cities using Dynamic Knowledge Graphs

Markus Hofmeister^{1,2,3}, George Brownbridge³, Michael Hillman³,
Sebastian Mosbach^{1,2,3}, Jethro Akroyd^{1,2,3}, Kok Foong Lee²,
Markus Kraft^{1,2,3,4,5}

released: June 6, 2023

¹ Department of Chemical Engineering
and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

² CARES
Cambridge Centre for Advanced
Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

³ CMCL Innovations
Sheraton House
Cambridge
CB3 0AX
United Kingdom

⁴ School of Chemical
and Biomedical Engineering
Nanyang Technological University
62 Nanyang Drive
Singapore, 637459

⁵ The Alan Turing Institute
2QR, John Dodson House
96 Euston Road
London, NW1 2DB
United Kingdom

Preprint No. 308



Keywords: knowledge graph, digital twin, smart city, flood assessment, interoperability

Edited by

Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

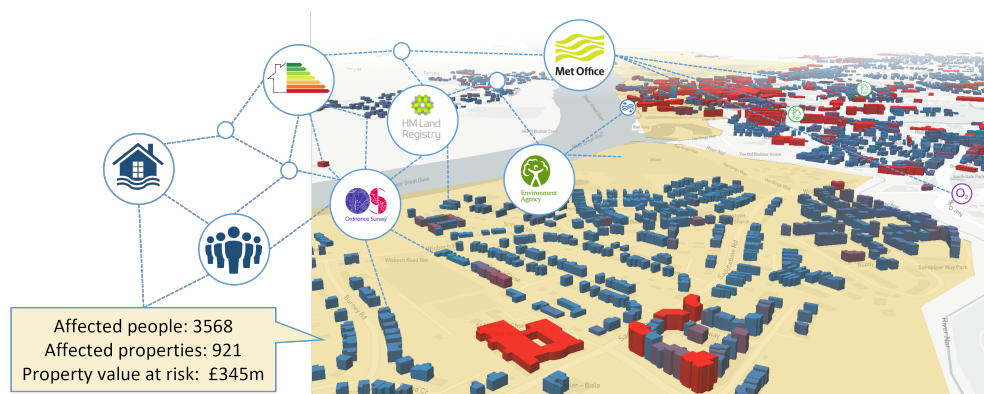
E-Mail: mk306@cam.ac.uk

World Wide Web: <https://como.ceb.cam.ac.uk/>



Abstract

This paper investigates the usage of knowledge graphs to bridge the gap between current data silos in deriving a holistic perspective on the impact of flooding. It builds on the idea of connected digital twins based on the World Avatar dynamic knowledge graph to deploy an ecosystem of autonomous software agents to continuously ingest new real-world information and operate on it. Multiple publicly available yet isolated data sources, including geospatial building information and property sales data as well as real-time river levels, weather observations, and flood warnings, are connected to instantiate a semantically rich ecosystem of knowledge, data, and computational capabilities to provide cross-domain insights in projected flooding events and their potential impact on population and built infrastructure. The extensibility of the proposed approach is highlighted by further integrating power, water, and telecoms infrastructure as part of the very same system, in order to analyse flood-induced asset failures and their propagation across networks. The World Avatar promotes evidence-based decision making during several disaster management phases, supporting both tactical and strategic risk assessments, which supports the United Nations Sustainable Development Goal 11 to improve the assessment of vulnerability, exposure, and risk of communities imposed by flooding events.



Highlights

- Dynamic integration of previously isolated yet related complementary data sources.
- Provision of holistic cross-domain perspective on potential flooding impacts.
- Fully automated updates of outdated information to ensure up-to-date insights.
- Single system to provide both tactical and strategic risk assessments.
- Unified visualisation interface to support evidence-based decision-making.

Contents

1	Introduction	3
2	Background	5
2.1	Existing Flood Tools	6
2.2	Ontologies and Knowledge Graphs	7
2.3	The World Avatar	8
3	Use Case	9
3.1	Public Data Sources	10
3.1.1	Building Data	10
3.1.2	Environmental Sensor Data	12
3.1.3	Flood Monitoring Data	13
3.1.4	Population Data	14
3.2	Knowledge Graph Instantiation	15
4	Results	16
4.1	Enabling Cross-domain Interoperability	16
4.2	Automated Cross-domain Flood Risk Assessment	20
4.3	Extensible system of systems approach	23
5	Conclusions	24
	Nomenclature	25
A	Appendix	27
	References	30

1 Introduction

Floods affect more people than any other type of natural disaster and the number of flood-related disasters is continuously increasing, endangering lives and leading to heavy economic losses [21, 39]. As climate change is likely to amplify climate variability, with extreme weather events becoming more severe and more frequent [39], especially devastating but short-duration urban floods are on the rise [43]. The United Nations, in its Sustainable Development Goals, has called for a transition from just responding to flood events to assessing the vulnerability, exposure, and risk of communities with the ultimate goal of mitigating potential devastating impacts. The European Union implemented a framework which requires all member states to identify and assess relevant flood areas, assess assets and humans at risk in these areas, and implement measures to reduce identified flood risk [21]. The UK Environment Agency's National Flood Risk Assessment shows that there are 2.4 million properties at risk of flooding from rivers and the sea in England, with further 2.8 million properties being susceptible to surface water flooding [19]. The expected annual damage to residential and non-residential properties at risk of flooding from rivers and the sea in England is estimated at more than £1bn. On top of that, a detailed "future flood map" of Britain, simulating the impact of flooding due to climate change, suggests that annual flood damage could increase by another 20% if carbon emission reduction pledges are not met [28].

Floods are very complex phenomena involving a large number of stakeholders and domain experts to collaborate seamlessly to ensure comprehensive risk analyses as well as efficient emergency planning and response [51]. A holistic assessment of floods and their potential impacts requires a cross-domain perspective, considering population, the built environment, network assets as well as their functional dependencies. Current challenges in urban flood management comprise multiple dimensions: improving the accuracy and timeliness of flood forecasting models by capturing localised and rapid changes in urban environments [25]; fostering coordination and collaboration among multiple stakeholders, including government agencies, emergency responders, urban planners, and the public; creating a comprehensive fact-basis for effective decision-making by increasing real-time monitoring of relevant parameters, such as rainfall intensity, water levels, and infrastructure conditions [22]; and ensuring adequate integration of data from different sources, systems, and stakeholders to maintain a comprehensive, aligned, and up-to-date overview of a flood hazard situation.

One of the coping strategies includes better flood predictions and earlier warnings based on increasing amounts of sensor data and advancing modelling capabilities [3, 10]. While this targets the real-time data availability constraint, this also amplifies potential interoperability issues due to heterogeneous types and formats. Multiple tailored platform solutions exist to consolidate flood related data and partially also assess impacts with regards to people and building stock at risk, such as FloodFactor [25], FloodMapp [10], or GIS solutions built on ArcGIS' Living Atlas of the World [16, 20]; however, all of them are quite limited in scope (*i.e.*, all target individual cities or regions within the USA) and are not easily extensible due to proprietary data models. Fragmentation and different schemas to represent data result in high friction when exchanging and integrating data between stakeholders and potentially lack the capability to disambiguate complex rela-

tionships across domains. Ontologies foster interoperability and resolve such ambiguities by providing aligned and unique descriptions for relevant concepts and their relationships. They help to avoid miscommunication between stakeholders stemming from naming inconsistencies across multiple domains and support context-aware information retrieval in interdisciplinary settings, such as flood management. Graph technology provides the opportunity to instantiate ontologies and connect data about various aspects of disaster management and response into a network of entities and their interconnections [34].

The World Avatar dynamic knowledge graph is designed as open system to foster interoperability and effectively address cross-domain questions, such as where flooding is occurring and whether it represents a risk [1]. It combines ontologies (*i.e.*, data definitions) with actual data instances (*i.e.*, from (open) APIs), and computational services operating on the instantiated data (*i.e.*, so-called agents). It follows a system of systems approach where individual task-oriented agents and their interplay help to describe a complex behaviour to gain comprehensive understanding and foster informed decisions. Multiple agents can be connected using the so-called Derived Information Framework [2] to ensure that newly instantiated information automatically traverses through the entire graph, including updates to all related information. This infrastructure can be utilised to dynamically assess the building stock value at risk whenever new flood information is ingested [33]. Moreover, several digital twins can work collectively to combine complementary perspectives and derive broader insights on flooding events. For example, the Climate Resilience Demonstrator digital twin, developed by the National Digital Twin programme, integrates asset and flood datasets, failure models, and system impact analyses to understand the effects of extreme weather, particularly flooding, on energy, water, and telecoms networks [15]. The demonstrator reveals how potential asset failures propagate both within but also across networks to enhance infrastructure resilience in the face of climate change.

The **purpose of this paper** is to demonstrate the capabilities of the World Avatar to dynamically ingest heterogeneous data from a variety of previously isolated sources to support a more holistic view on flooding. Sequences of autonomous and connected agents are deployed to automatically assess potential impacts of expected flooding events with regards to population and built infrastructure at risk. The versatility of this approach is demonstrated by combining both tactical short-term as well as strategic long-term perspectives in the very same system, highlighted by integrating scenario analyses of flood-induced asset failures and their propagation across networks into the dynamic evaluation of affected people and buildings.

The structure of this paper is as follows: Section 2 summarises existing flood tools and previous technical works; section 3 introduces the target use case, relevant data sources and details on the knowledge graph instantiation; section 4 highlights the results and gained insights from the knowledge graph-based digital twin; and section 5 concludes the work.

2 Background

Goal 11 of the UN Sustainable Development Goals [49] recognises disaster risk reduction as an integral part of social and economic development. Numerous studies have been conducted to create detailed understanding of flood risks and vulnerabilities to support strategic planning: Chang et al. [9] have created an urban flood vulnerability index using a social-ecological-technological systems vulnerability framework to examine the complexity of urban floods. A generic framework for multi-dimensional risk assessment has been proposed by Ekmekcioğlu et al. [18] to prioritise certain city districts regarding flood hazard, flood vulnerability, and flood risk based on a multi-tiered comprehensive decision-making procedure. Given the increasing importance of urban flash floods, Mohtar et al. [38] have developed an integrated flash flood index to assess susceptibility, vulnerability, and socio-economic impact of urban flash floods, in particular with regards to buildings and infrastructure. The main purpose of all these efforts is to identify critically flood-prone areas and provide assistance in improving resilience and mitigation plans.

Flood-induced damages to infrastructure networks can be costly, and are, compared to extensive research on impact on communities and economies, rather under-researched [41]. For example, in the 2007 summer floods in the UK, of the £4bn damage to the economy, approximately £670m was credited to damages to critical infrastructure [29]. The impact of compound flooding (*i.e.*, pluvial (rain), fluvial (rivers), and coastal flood) and its cascading effects on infrastructure systems in densely populated coastal areas has been studied by Najafi et al. [41]. While this work provides insights into potential failure cascades, it relies on expert judgement to create logical connections between individual assets and is limited to static analyses. The Climate Resilience Demonstrator (CReDo) investigates the impact of extreme weather, in particular flooding, on the energy, water and telecoms networks to identify weak spots and potential failure cascades to increase overall system resilience at lowest possible cost [15]. The CReDo digital twin uses information about the type, operational state, and location of each asset as well as the physical and logical connectivity between them to resolve the cascade of effects caused by a failure in any of the networks. The digital twin is based on the World Avatar (TWA) dynamic knowledge graph (KG), introduced in section 2.3, to combine a description of the assets with data from flood simulations for different climate change scenarios. The use of KG technology ensures the creation of an extensible ecosystem for connected digital twins that support the interoperability of distributed data across sectors, and ensures that data are connected, discoverable and queryable via a uniform interface [1].

Flood impacts vary with time. In the short term, they pose risks to life, cause property damage, and lead to infrastructure failures (*e.g.*, transport and electricity networks). In the medium term, contaminated flood waters increase the risk of the spread of diseases. In the long term, flooding-induced disruptions can have economic consequences beyond the immediately affected region [29]. Much research has been conducted in the area of short-term flood impact assessment, either using simulated flood scenarios or the post-evaluation of actual flood events [3, 29, 44]. Flood impacts can be categorised as tangible (property damage or financial losses) and intangible (loss of life, environmental impacts, *etc.*). Another distinction is between direct impacts (resulting from immediate physical contact with floodwater) and indirect damage (induced by direct impacts, *e.g.*, stock price

reduction). Direct tangible damage represents the best understood class of flood impacts and encompasses physical harm to properties, contents, and infrastructure due to direct exposure to floodwaters [29]; however, the majority of the works focuses on static evaluations and scenario planning instead of providing a live view on value at risk for a current flood hazard.

Credible estimates of flood impacts are important to provide decision support and allow for efficient resource allocation and risk management [29, 47]. The assessment may cover environmental, economic, and social perspectives, or all three. To integrate the different perspectives, either a common metric can be applied (*i.e.*, almost always in monetary terms), or impacts can be assessed using multi-criteria techniques. Flood impacts on buildings, infrastructure, and land-use are commonly assessed using damage functions that link expected damage to flood characteristics such as extent, depth and flow velocity [3, 44, 47]. Damage functions exist in various complexities and often distinguish between different asset types. It has been shown that the influence of different damage models is small compared to the influence of different hazard and exposure maps used to identify flood-affected objects [47].

2.1 Existing Flood Tools

The European Flood Awareness System offers an early warning system for real-time monitoring and forecasting of floods across Europe, aiding in flood preparedness and response. It supports preparatory measures for anticipated flood events; however, data access is limited to partners and is not publicly available [22]. The UK, however, provides a rich body of data enabling the creation of high-resolution, high-accuracy flood hazard datasets to provide indications of the magnitude and probability of flooding. While several publicly available flood assessment tools exist to provide some consolidated insights, most of them remain limited to individual domains. For example, a live flood map [37] shows current flood warnings, together with current readings for river, sea, groundwater, and rainfall levels, and the expected flood risk over the next 5 days. Additionally, separate governmental services are available to raise notifications for all registered users in the vicinity of issued flood warnings. However, none of the available tools provides real cross-domain insights, such as assessing the impacts of a potential flood in terms of people and built infrastructure at risk. Our work addresses this gap, partially inspired by related efforts in the United States, such as FloodFactor [25] or FloodMapp [10].

FloodFactor [25] provides a free online tool that equips individuals with the ability to understand whether a property has flooded in the past or is currently at risk, and how that risk is likely to change over time. It supports estimating damage cost associated with flooding and highlights infrastructure and community risks. It relies on a comprehensive flood model initially developed by Bates et al. [3], which analyses flood hazards (including pluvial, fluvial, tidal events, and storm surge), projects future climate scenarios, incorporates local adaptation, and validates against modeled historic floods as well as satellite images and government records. Originally aimed to provide the general public with insights into the effects of flooding events on property values (*i.e.*, to level the playing field compared to institutional investors, and promote greater awareness of flooding within society), FloodFactor has evolved into one of the most precise probabilistic flood models available

in the United States. It includes the analysis of flood risk for social facilities, critical infrastructure, commercial properties, and roads; however, it does not provide insights into failure cascading and is rather focused on static scenario analyses, such as estimating the total share of infrastructure assets at risk.

FloodMapp [10] provides real-time flood intelligence tools for emergency managers, with tailored products for each stage of the disaster management cycle: flood forecast live mapping up to seven days before a flood event to plan evacuations and protect sites and assets (updated hourly); real-time flood inundation extents and depths mapping during a flood event to maintain situational awareness and take targeted actions (updated every 15 minutes); and detailed flood depth and inundation maps immediately after a flood event to rapidly assess damage and initiate recovery. The service is not open data and currently limited to the United States and Australia, where real-time and forecast rainfall, tidal and river height data are used together with machine learning techniques to create a dynamic 1 m resolution flood inundation and extent model. However, the focus is clearly on providing dynamic high-resolution flood mapping instead of up-to-date impact assessments.

2.2 Ontologies and Knowledge Graphs

Ontologies are formal representations of knowledge capturing concepts, relationships, and properties within a specific domain. They provide a structured framework for organising and sharing information in a machine-readable format. Representing data using ontologies results in the formation of directed graphs, so-called knowledge graphs (KGs), where nodes define entities and data (*i.e.*, concepts or instances) and edges denote their relationships [1]. KGs provide an extensible data structure that is well suited to represent arbitrarily structured data and which can be hosted decentralised, *i.e.*, distributed over the internet, using Semantic Web technology [4]. This takes the form of Linked Data [5], where every concept and relation can be referenced back to its definition, making information discoverable across the web and providing additional context information to enable machine readability and automation.

Linked Data supports FAIR data principles [50] which help overcome interoperability issues due to information silos, improve data clarity and resolve inconsistencies. Linked Data enables the interconnection of diverse datasets by utilising subject-predicate-object triples, where the subject represents a resource, the predicate indicates a specific property or relationship, and the object represents a value or another resource. In the context of flood risk, an ontology could define concepts like `River`, `WaterLevel`, and `Damage`, while Linked Data could connect these concepts through subject-predicate-object triples, such as `River A - has - WaterLevel X` or `River A - causes - Damage Y`.

The use of KGs has gained traction as a vital technology for offering machine-interpretable, semantic information about real-world entities on a large scale. For example, Johnson et al. [34] have presented a scalable workflow for merging data from OpenStreetMap, Microsoft USA building footprint layer, and the OpenAddress project to create a comprehensive KG of urban infrastructure data. Machine learning models are applied to the KG to infer and populate potential gaps in order to ensure availability of important information for emergency responders in case of a flood event. GeoSPARQL queries are

used to assess both built infrastructure and demographics at risk of flooding; however, no dynamic data assimilation is supported and a new KG needs to be created on demand for each new analysis.

2.3 The World Avatar

The World Avatar project [1] aims to create a digital ‘avatar’ of the world. The fundamental idea behind TWA is the creation of an all-encompassing world model to foster interoperability between previously isolated but conceptually connected domains, both in terms of knowledge and data. Initially centered around chemical and process engineering [35, 52], TWA has meanwhile expanded to address decarbonisation challenges in the energy sector [14, 45, 46] and overcome interoperability constraints during city planning [6–8]. Consistent semantic descriptions of relevant concepts as well as their relationships are provided by ontologies. The digital world is represented using a dynamic KG that contains concepts and data that describe the world together with an ecosystem of autonomous computational agents to simulate its behaviour. Computational agents are software tools which are described ontologically and are themselves parts of the KG. This allows computational capabilities to become discoverable themselves, which means that the graph provides information about its data as well as what could be done with it. A conceptual illustration of TWA is shown in **Fig. 1**.

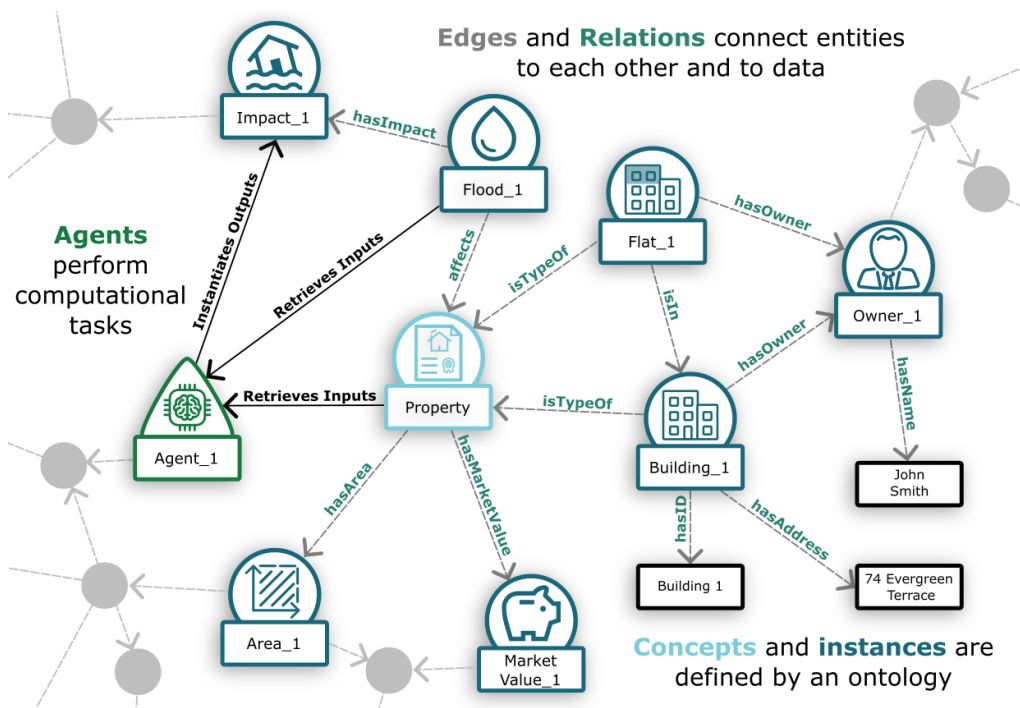


Figure 1: The World Avatar dynamic knowledge graph. Representing data using ontologies results in the formation of directed graphs, where nodes define entities and data (i.e., concepts or instances) and edges denote their relationships. Computational capabilities to operate on the data, so-called agents, are an integral part of the graph making it inherently dynamic.

TWA is modular and scalable by design, supporting both decentralisation and interoperability across heterogeneous data sources and software. The intelligent agents integrated within TWA function as executable knowledge components, ensuring the system remains up-to-date and self-evolving. TWA offers utility in three key aspects: 1) providing cross-domain insights into the current state of the world, 2) controlling real-world entities, and 3) facilitating complex what-if scenario analyses. As everything is connected (*i.e.*, data, concepts, and agents), this design enables a system of systems approach to create an ecosystem of connected digital twins. Compared to platform solutions, TWA promotes open protocols and standards to avoid potential lock-in effects and surging switching cost.

The Derived Information Framework [2] is a knowledge-graph-native solution for tracking data dependencies and managing information flow within TWA. It is focused on data provenance, aiming to identify the source of information and how it has been obtained. By representing intrinsic dependencies within the KG, it enables autonomous data handling by agents with the ability of information to cascade automatically through the KG. In flooding contexts, this capability can be deployed to automatically assess potential impacts on population and building stock: Multiple input agents collect (real-time) data from various publicly available sources (detailed in section 3.1) to create an accurate representation of the built environment as well as environmental conditions of the real world within TWA. Subsequently, three autonomous agents, connected through the Derived Information Framework, ensure that all dependent information are updated in conjunction with newly ingested data. Consequently, whenever new flood warnings are raised and assimilated, up-to-date estimates of the total building stock value and population at risk are available. While the technical details are discussed elsewhere [33], this work focuses on the gained insights and the added value for flood risk management.

3 Use Case

This work aims to address the urgent need to reduce imminent risk and impact of flood hazards along three dimensions. While the approach is generally applicable to the entire UK, this proof of concept focuses on King’s Lynn, a mid-size coastal town in the East of England:

Firstly, flood related data from numerous sources is instantiated using consistent knowledge models, enabling interoperability between previously isolated information and creating a more holistic perspective to evaluate flood risk. Connecting complementary data enables both humans and software agents to make better fact-based decisions, *e.g.*, by bringing together data about potential flood events (*i.e.*, severity, areal extent) with data about the built environment (*i.e.*, building locations, building usages, property values), or allowing plausibility checks of sensor readings across multiple sources (*e.g.*, verifying rising river levels with precipitation data from another source). A uniform query and visualisation interface for the connected dataset is provided to ease access to cross-domain insights about (imminent) flood events (*e.g.*, understand which building types are mostly affected).

Secondly, the Environment Agency (EA) frequently issues alerts and warnings for potentially hazardous flood situations; however, such flood warnings lack information about

anticipated impacts on people and buildings. Our infrastructure addresses this gap and facilitates the dynamic integration of new data as well as automated (re-)assessments of all dependent information within the KG. This includes automatic impact re-estimations whenever flood warnings are issued or any other relevant input data gets updated, providing an up-to-date view of the situation at all times.

Thirdly, the very same infrastructure also supports strategic analyses beyond the dynamic assessment of immediate flood impacts. TWA can be used to understand how flood-induced infrastructure failures cascade across different utility networks to identify critical nodes and potential weak points. Together with strategic flood risk maps this can help to increase overall flood resilience.

3.1 Public Data Sources

A thorough search for open data sources related to built infrastructure, environmental measurements, and flood data has been conducted, with a primary focus on the vicinity of King’s Lynn. Several static sources and application programming interfaces (APIs) have been identified and are summarised below. **Tables 1** and **2** offer an overview of the considered data feeds, with additional details provided below. For a comprehensive discussion on data instantiation, please refer to Hofmeister et al. [33].

3.1.1 Building Data

The Ordnance Survey (OS) is the national mapping agency of Great Britain responsible for producing and disseminating geospatial data for government and public use [42]. It provides several open (*e.g.*, OpenMap Local) and premium (*i.e.*, Building Height At-

Table 1: *Publicly available (rather) static data sources considered for instantiation of building data.*

Data source	Instantiated parameters	Update frequency
Ordnance Survey (via Digimap) [42]	Building footprint Building elevation Building height UPRN	Initial download and static instantiation
Energy Performance Certificate API [13]	Property type and built form Property usage category Address (incl. postcode) Number of habitable rooms Total floor area Energy rating UPRN	Every 4-6 months
His Majesty’s Land Registry [30]	Property sales transaction record (<i>i.e.</i> , address, sales price, date) UK House Price Index	Monthly

tribute) datasets describing the physical characteristics of the built environment in various levels of detail. While the Building Height Attribute (BHA) data represents the most granular data about individual buildings, including their base polygons, building heights, and ground elevation, the OpenMap Local contains building data on a more aggregated level and lacks information about building heights. Premium datasets are generally license restricted; however, made available via Digimap [17] for educational and research purposes. It has been observed that some buildings are only contained in either of the two datasets. Therefore, both the BHA and OpenMap Local data have been geospatially merged to create the most comprehensive building coverage possible. Wherever building information from both datasets is available, the BHA data is used. Consequently, the geospatial building information used in this study is mainly based on the OS BHA dataset. Both datasets contain the Unique Property Reference Number (UPRN) for each building, which constitutes a unique and officially maintained identifier assigned to every addressable location in the UK (*i.e.*, used to cross-link information across datasets).

The Department for Levelling Up, Housing & Communities offers open Energy Performance Certificate (EPC) data [13] via three dedicated APIs for domestic, non-domestic and display (*i.e.*, mostly public buildings) certificates. This data contains property-level information about energy efficiency rating, environmental impact rating, and recommendations for improving energy efficiency. Additionally, it provides details about a building's heating, cooling, and ventilation systems, key construction characteristics (*i.e.*, number of rooms, total floor area, building type, *etc.*), high-level usage classification, address and location details, and is updated every four to six months. The EPC data can be used to enrich the sole geospatial building representation from OS (*i.e.*, via UPRN matching) with key features required to conduct energy analyses or estimate property prices. Available data comprises EPCs issued for domestic and non-domestic buildings constructed, sold or let since 2008 and contains approximately 60% of the housing stock in England. Despite a quite homogeneous coverage across all regions, the data should not be interpreted as a complete representation of the building stock in England, since not every single building is included.

His Majesty's Land Registry publishes several public datasets related to residential property sales, namely the Price Paid Data and the UK House Price Index (UKHPI) [30]. The Price Paid Data tracks residential property sales in England and Wales, with new transactions being added every month [31]. Data is captured for single residential properties that have been sold and lodged with Land Registry since 1995, making it a reliable source of house price information with currently more than 24 million records. The recorded data contains actual prices and transaction dates, together with full address details and the type of property, such as detached, semi-detached, terraced or flat. Since UPRN information is not included in the recorded data, previous property sales transaction records can be combined with further building information via address matching only. The UKHPI captures the monthly change in the value of residential properties (*i.e.*, with regards to a base of 100 set to January 2015) on different levels of geospatial granularity [32] (details in Appendix A.1).

3.1.2 Environmental Sensor Data

The UK Meteorological Office (abbreviated as Met Office) provides a collection of near real-time observation and forecast weather information through its public API known as DataPoint [36]. DataPoint includes observations for approximately 140 sites and forecasts for approximately 6,000 sites across the UK. Actual weather observations are provided with hourly resolution for the past 24 hours, while forecasts are created with daily and three-hourly resolution for the next five days. Both data feeds get updated hourly. Observation reports are recorded in real time by the Met Office Monitoring System, and reported parameters depend on the physical instrumentation installed at each site. It needs to be noted that the published observation data has not yet been subject to final quality control by the Met Office. Table 2 provides an overview of all available weather parameters.

Table 2: Publicly available near real-time data sources considered for instantiation of environmental observation data.

Data source	Available parameters	Data resolution	
		Spatial	Temporal
UK Met Office DataPoint [36]	Temperature Dew point Pressure Relative humidity Visibility Wind speed Wind gust Wind direction Precipitation probability UV index	~140 observation and ~6000 forecast stations across the UK	Observations with hourly resolution (past 24 h), Forecasts with 3h resolu- tion (next five days), both updated hourly
UK-AIR Sensor Observation Service [48]	Nitrogen dioxide Nitrogen monoxide Sulphur dioxide Carbon monoxide Ozone PM _{2.5} PM ₁₀	~300 observation locations across the UK (organised in different networks; not all locally- managed networks published via API)	Observations with hourly resolution (including his- torical data)
River Levels ¹	Water level River flow rate Rainfall	~4000 river stations and ~1000 rainfall gauges across the UK	Observations with 15 min resolution (updated 1-2 times per day)
Flood alerts/ warnings ¹	Affected flood area Severity Warning description Warning change log (<i>i.e.</i> , timestamps of changes)	Entire UK, covering specific warning and alert areas	Updates provides every 15 minutes

¹ Provided via EA Real Time flood-monitoring API [11]

The UK Air Information Resource (UK-AIR) collects data from a network of air quality sensors across the UK, providing real-time and historical information on pollutants,

such as nitrogen and sulphur oxides, particulate matter (*i.e.*, PM_{2.5} and PM₁₀), and ozone, via an open machine-readable Sensor Observation Service (SOS) [48]. There are two types of monitoring networks: automatic and non-automatic. While automatic networks provide hourly pollutant concentrations collected from individual sites using modems, non-automatic networks measure less frequently (*i.e.*, daily, weekly, or monthly) using physical collection methods like diffusion tubes or filters. There are around 300 EA managed monitoring sites across the UK as part of a national monitoring strategy. Those sites are organised into several networks focusing on particular observation tasks, and further complemented with locally-managed monitoring networks, including sites operated by local authorities, industry partners, and airports. The location and purpose of these locally-managed sites differ from the national network and are subject to different regulations. The data from these sites are stored in various databases operated by different custodians and are collected from publicly available online resources or provided to UK AIR voluntarily via a dedicated API. UK-AIR provides an interactive monitoring map to access the data from both national and locally run networks: The map includes data from the majority of automatic network stations as well as some non-automatic ones. Additionally, approximately 80% of the automatic sites from local authorities are available on the map. However, it should be noted that not all local networks, for example King's Lynn and West Norfolk, are available via the SOS.

3.1.3 Flood Monitoring Data

The EA offers several API endpoints with (near) real-time information related to flooding and flood risk: The Real Time flood-monitoring API [11] provides a listing of all current flood alerts and warnings as well as information about flood areas to which those apply, including relevant meta information, such as severity and associated water bodies. The same API also includes an endpoint for live readings of water levels and flows recorded at various measuring stations along rivers and other water bodies. Hydrological data about river levels, river flows, groundwater levels, and water quality is provided via EA's Hydrology API [12]. Moreover, the Flood Forecasting Centre issues a machine-readable daily flood risk forecast, assessing the likelihood of flooding 5 days into the future [26]. All APIs are provided as open data under the Open Government Licence.

River Levels and Rainfall The flood-monitoring API provides measurements of water levels and flow rates as well as information on the monitoring stations conducting those measurements. Water levels and flow rates are usually measured every 15 minutes. However, the actual data transfer frequency depends on the monitoring station and the current level of flood risk, but typically occurs once or twice per day and usually increases during times of heightened flood risk. The latest release of this API, furthermore, provides access to approximately 1,000 real-time rain gauges [11], which measure the amount of precipitation (mm) using tipping bucket gauges. Due to data protection reasons the geographic location of the rainfall monitoring stations has been reduced to a 100m grid.

The Hydrology API complements the near real-time water level and rainfall information from the flood-monitoring API by providing access to historical water level and flow rate information [12]. The data model differs slightly between the two sources and the

hydrology data includes flags to indicate the quality of the provided data, as its majority is subject to (manual) quality control before publishing.

Flood Alerts and Warnings The EA issues a list of currently active flood warnings for specific flood warning or alert areas, which is updated every 15 minutes. The provided information for each flood alert or warning includes its severity classification, associated water bodies, the source of flooding, a continuously updated warning message, and the flood area as GeoJSON polygon. Flood alert areas denote geographical areas where it is possible for flooding to occur from rivers, sea, or groundwater. A single flood alert area may contain a number of flood warning areas. Flood warning areas are geographical areas where the EA expects flooding to occur and where the flood warning service is provided. Specifically, those flood warning areas can specify discrete communities at risk of flooding. There are four distinct severity levels of decreasing magnitude: Level 1 represents a *severe flood warning* with potential danger to life or widespread disruption. Level 2 refers to expected flooding with immediate actions to be taken. The purpose of such a *flood warning* is to alert people that flooding is expected and they should take action to protect themselves and their property. Level 3 denotes a *flood alert* issued to warn people of the possibility of flooding and encourage them to be alert and make early preparations. Severity level 4 indicates that a previous flood warning is no longer in force. While much of the information is provided using stable long term Uniform Resource Identifiers (URIs), such as flood areas or measurement stations, some resources only exist temporarily. For example, the URI of an individual flood alert will only exist and resolve successfully while the alert is in force. Both severity and warning message associated with an alert can change during its lifetime. And eventually, any alert or warning will cease and its URI will no longer resolve.

Flood guidance statements issued by the Flood Forecasting Centre represent the best combined understanding of flood risk based on weather forecasts, flood forecasts, catchment conditions, and the operational status of flood defences [27]. The public five day forecast [26] shows the high level content of a flood guidance statement, appropriate for a public audience. Compared to immediate flood alerts and warnings, flood forecasts are associated with higher uncertainty, both with regards to areal extent and anticipated severity.

3.1.4 Population Data

The OpenPopGrid [40] provides an open and high resolution (*i.e.*, 10m x 10m) gridded population dataset for England and Wales based on the Office for National Statistics (ONS) 2011 Census as well as OS OpenData. It is specifically designed to enhance the spatial representation of the published ONS population data, which is originally provided for so-called Census Output Areas based on aggregated unit postcodes. Using a dasy-metric mapping approach, the population is redistributed to residential areas based on building polygons in order to derive a realistic depiction of where people actually live.

To address potential issues regarding the age of the dataset, a comparison with the continuously updated HDX population map [24] maintained by Meta has been conducted.

The HDX population map incorporates various sources like census data and surveys, and also considers anonymised Facebook usage data to estimate population density at a high resolution. It is commonly used in public health research and humanitarian efforts. The HDX dataset is slightly coarser; however, both datasets align well on the Super Output Area Middle Layer (*i.e.*, a typical Census Output Area), with the OpenPopGrid providing approximately 4% lower population estimates (details in Appendix A.2). Given this small discrepancy, the OpenPopGrid data is still considered representative and used in this study due to its greater geospatial resolution and expressiveness of where people live.

3.2 Knowledge Graph Instantiation

All relevant data described above are instantiated by an ecosystem of software agents as described in Hofmeister et al. [33]. All agents are deployed collectively in the cloud to ensure reliable service availability. As this use case focuses on the automated cross-domain flood impact assessment, both the Hydrology and Flood Forecast APIs briefly introduced above are not instantiated. A few further limitations and simplifications need to be noted:

Firstly, only buildings with a geospatial representation are considered in this analysis, as the geolocation is required to understand which buildings are affected by a certain flood warning. Hence, EPC and property sales data is only instantiated for buildings with instantiated geospatial representation, using UPRN and address matching, respectively. While more than 37,700 buildings are instantiated with regards to their geometry, only about 21,200 (56% of geospatially represented buildings) of them are assigned UPRN information. This likely stems from unavailable UPRN information in the original OS data (details in Appendix A.3) and appears to affect all building types and built forms equally. Since no significant distortion could be identified in the distribution of instantiated building stock, the available buildings are considered a representative subset and used as baseline for further building data enrichment. Based on the instantiated UPRNs, approximately 14,700 properties could be linked to EPC information, of which approximately 11,000 (29% of geospatially represented buildings) correspond to buildings (*i.e.*, remaining portion referring to individual flats).

Secondly, the current approach to estimate the market value of properties is rather simplistic. It involves scaling the latest historical transaction record based on the UKHPI or using average square metre prices for the respective postcode. However, in reality, numerous factors influence real estate values, including local market conditions, property type and age, number of rooms, nearby infrastructure, *etc.*

Thirdly, the air quality monitoring in the vicinity of King’s Lynn belongs to a locally-managed monitoring network not (yet) included in the UK-AIR SOS. Hence, no real-time data can be retrieved automatically via the API. To still demonstrate the general capability, a virtual station is instantiated in the town of King’s Lynn, showing actual data from the nearest available monitoring site.

4 Results

This section describes the results of the proof of concept implementation: In subsection 4.1 we show how TWA effectively connects previously isolated but related information to foster more holistic and fact-based decision-making. Data is connected based on aligned knowledge models and visualised in an uniform web interface. The automatic assessment of potential flood impact based on automatic information cascading is presented in subsection 4.2. Subsection 4.3 exemplifies the extensibility of TWA by seamlessly integrating with the more strategic CReDo visualisation, facilitating comprehensive short- and long-term analyses within one unified system.

4.1 Enabling Cross-domain Interoperability

The World Avatar acts as single point of access to current weather, air quality, and flood related information. In the focus area of this use case (*i.e.*, the vicinity of King’s Lynn), additional data about the built environment, including building usages, location, construction details, and previous property sales transactions, are included. The developed agent framework aligns and enriches this data by introducing links between related entities in order to add a geospatial dimension to property sales data (*i.e.*, by combining it with OS and EPC data via address matching) and using automated agent tasks to derive up-to-date property market value estimates for all instantiated properties. TWA’s unified web-based visualisation interface presents this comprehensive information in a map-based format, accessible through standard web browsers. **Fig. 2** provides an illustrative example of the visualisation.

A consolidated visualisation that bridges the gap between different information silos empowers residents of smart cities (as well as software agents) by increasing transparency and ease of access to a wide variety of available data. For example, individuals can easily access the environmental observations most relevant to them, such as air quality or river water levels closest to their home. While such a visualisation enhances user experience and promotes data-driven decision-making, the actual key benefit of the KG instantiation is the creation of semantically linked datasets: Meaningful links between related data provide further context information, *e.g.*, by linking air pollutant observations to equivalent concepts of the European air quality e-Reporting initiative [23] in order to obtain current pollutant limits and protection targets. Furthermore, direct inference can be used to assess whether an observation from a sensor should trigger an alert or any further action.

The unified visualisation interface enables users to simultaneously view various flood-related data, such as the latest water levels, rainfall readings, weather forecasts, and active flood warnings. The consolidated visualisation provides broader context for individual information and helps stakeholders to assess potential flood hazards more holistically (see examples provided in **Fig. 3** and **Fig. 4**). Additionally, it facilitates cross-domain reasoning and enables data plausibility checks across multiple data providers.

Figure 3 presents an actual situation in Cambridge on the evening of 09 March 2023: At 5pm, a flood warning was issued as a precautionary measure due to rising river levels caused by heavy rainfall and snow melt. The forecast indicated that rainfall is expected

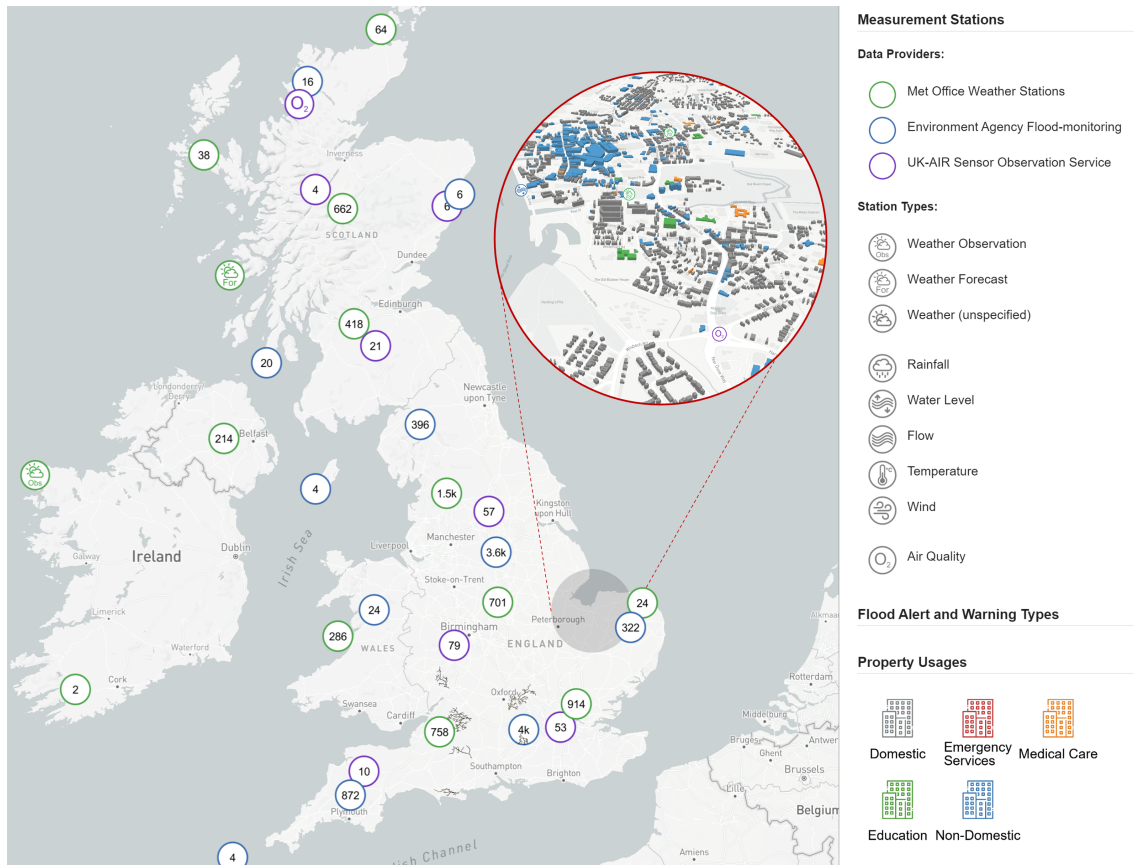
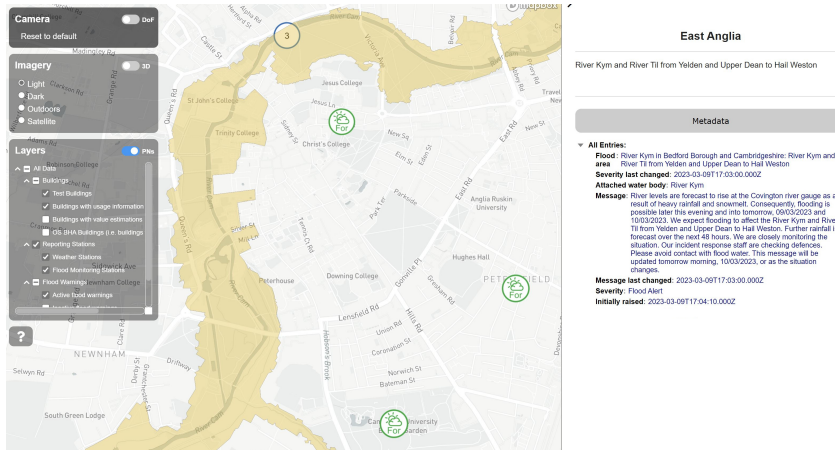
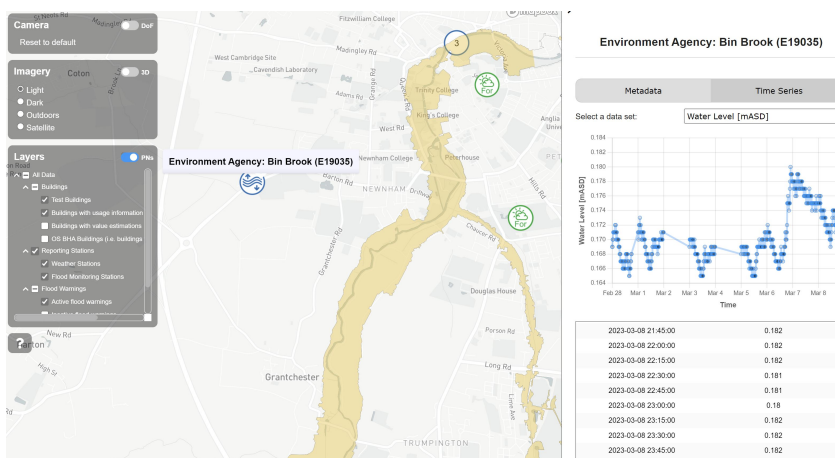


Figure 2: *The World Avatar provides consolidated and scalable access to previously isolated environmental observations, including weather, flooding, and air quality data. Information about the built environment, such as buildings, building usages, and property value estimates, complement this perspective.*

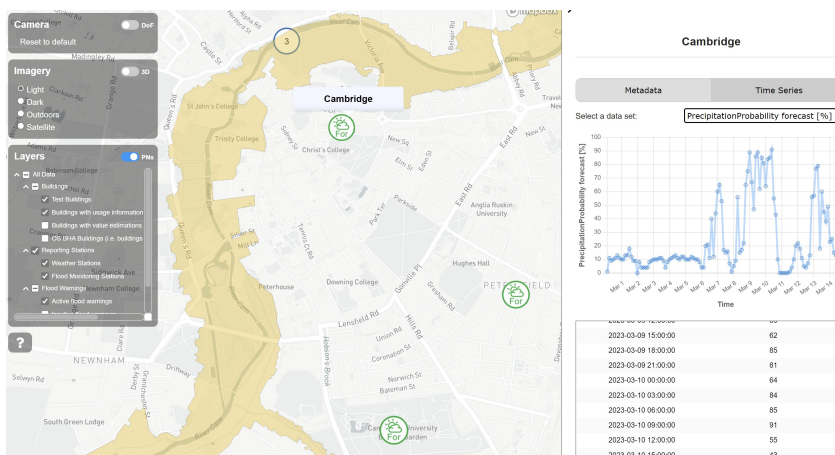
to continue for the next 48 hours, which increased the likelihood of flooding on both 09 March and 10 March. These predictions were aligned with the live water level readings of affected rivers, providing additional confirmation. Additionally, the forecast precipitation probability by the Met Office supported the warning for ongoing rainfall; however, also suggested a period of relief as of the second half of 10 March. Figure 4 illustrates the actual situation in Christchurch Harbour in the afternoon of 08 March 2023: A flood alert was raised due to the anticipated spring tides, which were influenced by strong south-westerly winds. The expected tidal surge posed a risk of overtopping the sea defenses, potentially leading to flooding of roads, farmland, and properties. The forecast spring tide levels of 1.31 and 1.34 m aligned closely with the latest observations from a nearby river level station. Furthermore, the forecast wind speeds for 08/09 March from Met Office supported the flood warning, although winds were rather expected to be on the higher side of Force 4 (13-18 mi/h) on 08 March.



(a) EA Flood Warnings API. A flood warning was issued at 5pm as a precautionary measure due to heavy rainfall, which was anticipated to persist and result in flooding.

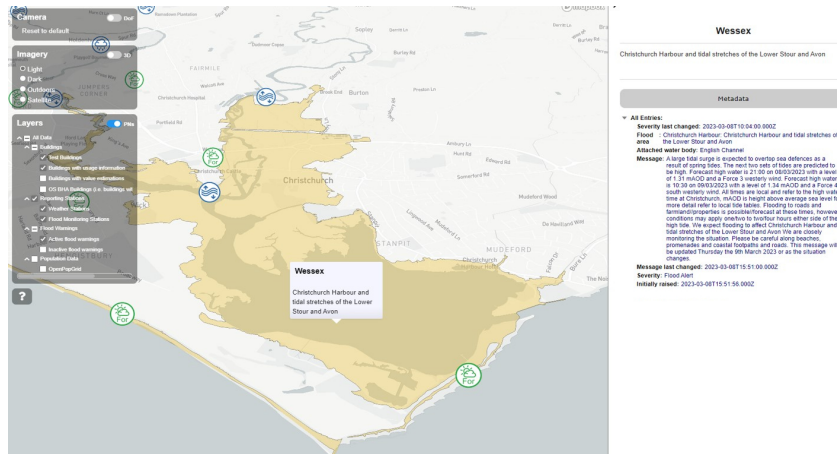


(b) EA River Levels API. Latest water level readings of affected rivers supported this judgement. (Please note: EA API faces infrequent periods of missing data)

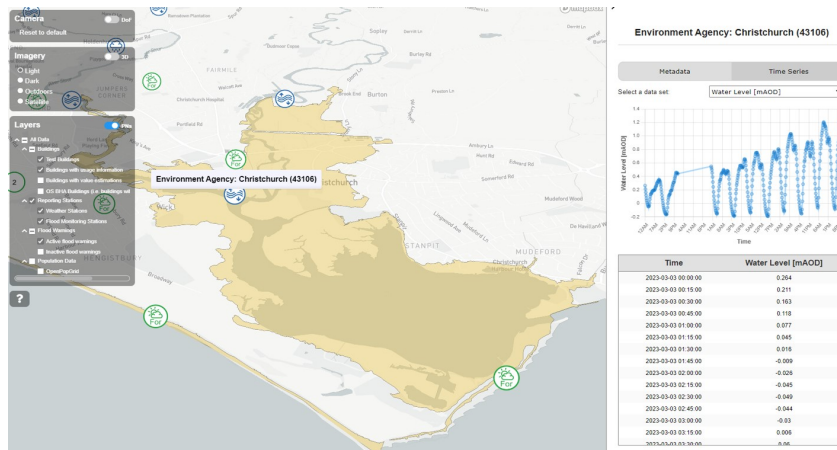


(c) Met Office API. Anticipated precipitation probability provided further evidence for possible flooding, but also indicated a period of relief thereafter.

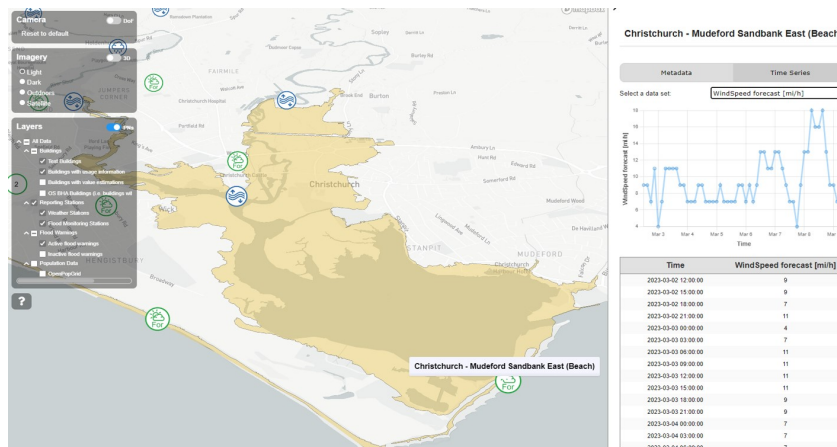
Figure 3: Actual situation for Cambridge in the evening of 09 March 2023.



(a) EA Flood Warnings API. Flood alert was raised due to expected spring tides, partially as a result of strong (south) westerly winds.



(b) EA River Levels API. Anticipated spring tide levels of 1.31 and 1.34 m were in line with latest observations at nearby river level station. (Please note: EA API faces infrequent periods of missing data)



(c) Met Office API. Forecast wind speeds for 08/09 March supported anticipated exacerbating effect on upcoming two spring tides.

Figure 4: Actual situation for Christchurch Harbour in the afternoon of 08 March 2023.

4.2 Automated Cross-domain Flood Risk Assessment

The World Avatar continuously assimilates latest buildings and flood warnings information into the dynamic KG. In conjunction with the automatic re-assessment of potential flood impacts, this ensures an up-to-date view on the real-world situation at all times. The impact estimation considers the number of people at risk, the number of potentially affected buildings, and their respective monetary value. Key inputs to the assessment include the flood alerts/warning itself (*i.e.*, severity, message, classification), instantiated buildings (*i.e.*, newly instantiated buildings, or updated floor area or sales transaction data), and the property price index. Any changes in these inputs are automatically reflected during the next flood impact assessment. The property price index is used to re-calibrate historical sales prices of properties to their present-day values. In cases of properties without previous sales transactions, an average square metre price per postcode is multiplied with the total floor area of the building to derive a current market value estimate. Hence, a re-assessment may occur even if none of the directly affected properties have been updated, but new property information for the postcode has become available to refine the previous average square metre price.

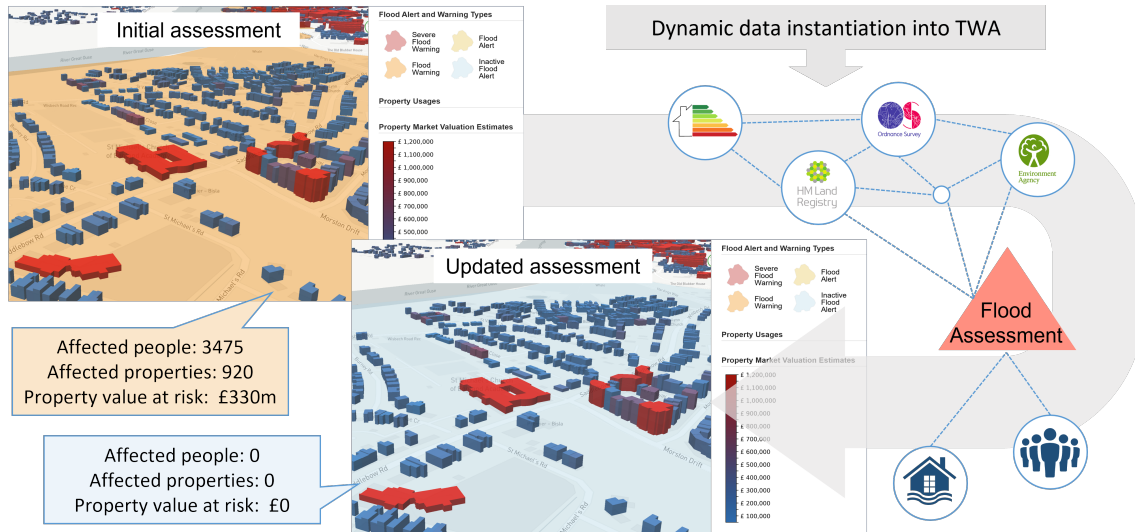
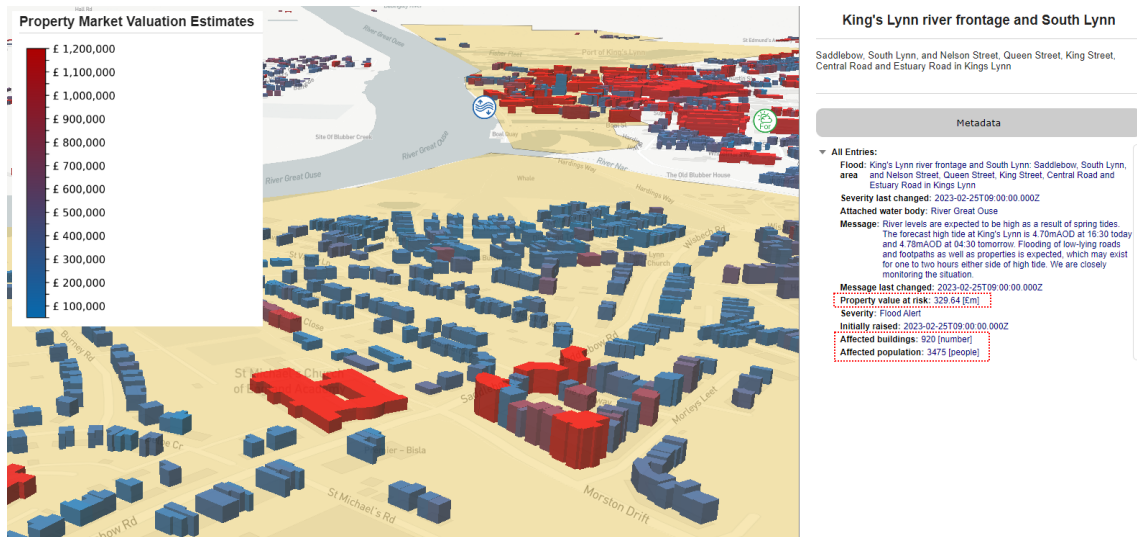
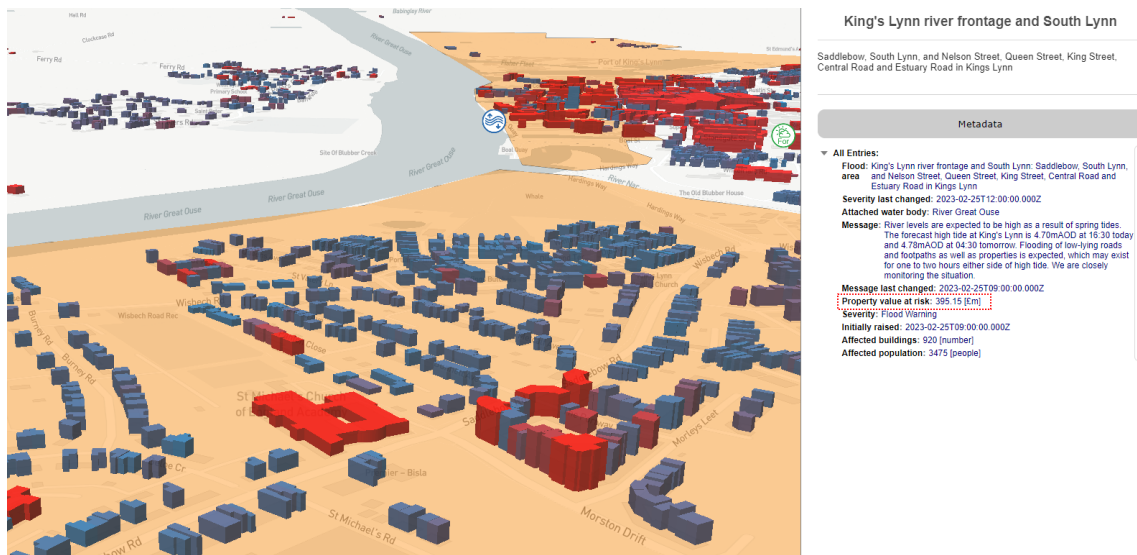


Figure 5: Automatic re-assessment of potential flood impact. Whenever relevant inputs for the flood assessment get updated (*i.e.*, flood warnings change or buildings data in vicinity, incl. respective property sales transactions, are updated), an automatic re-assessment of the potential flood impact is triggered. Illustrated for updated flood warning severity, where a previously active warning ceases and is finally lifted, resulting in impact assessed as zero.

The agent based re-assessment is schematically illustrated in **Fig. 5** for the case of a ceasing flood hazard. Real-world updates, including changes in flood warning severity, are instantiated into TWA. The Flood Assessment Agent automatically detects that its previous assessment has become outdated and performs a re-evaluation accordingly. In the case of ceasing flood warnings, the agent sets the number of affected people and buildings as well as the property value at risk to zero, indicating the absence of imminent flood risk.



(a) Initial flood impact assessment. Based on latest real-world data the raised flood alert puts 3475 people at risk and could affect 920 buildings with an estimated market value of £329.6m.



(b) Updated flood impact assessment. To simulate changes in latest property prices, the underlying property price index is scaled by +20%. The instantiation of an updated flood warning triggers an automatic re-assessment, which is now expected to affect (still) 3475 people as well as 920 buildings; however, the estimated market value now sums up to £395.2m.

Figure 6: Automatic re-assessment of potential flood impacts for updated property values, i.e., newly instantiated property sales transactions and/or updated property price index.

Similarly, any changes in property prices (i.e., due to updated property price index data or newly instantiated transaction records) will result in an updated impact assessment, as shown in **Fig. 6**. The figure shows two subsequent evaluations side by side: the initial estimation following a newly raised flood alert, and the updated assessment after an increase in property prices. The initial assessment identifies 3,475 people and 920 buildings at risk,

with an approximate market value of £ 330mn (Fig. 6(a)). After the initial assessment, a 20% increase in the property price index is instantiated to simulate a rise in the value of residential properties. Subsequently, the Flood Assessment Agent automatically detects the need for re-evaluation due to the change in property prices. Although the number of people and buildings at risk remains the same, the total property value at risk increases to £ 395mn (Fig. 6(b)). It is important to note that this increase is for illustrative purposes and is not based on specific market dynamics. However, this scenario demonstrates the system’s dynamic capabilities, *e.g.*, to automatically reflect adjustments in the monthly UKHPI in any subsequent flood assessment.

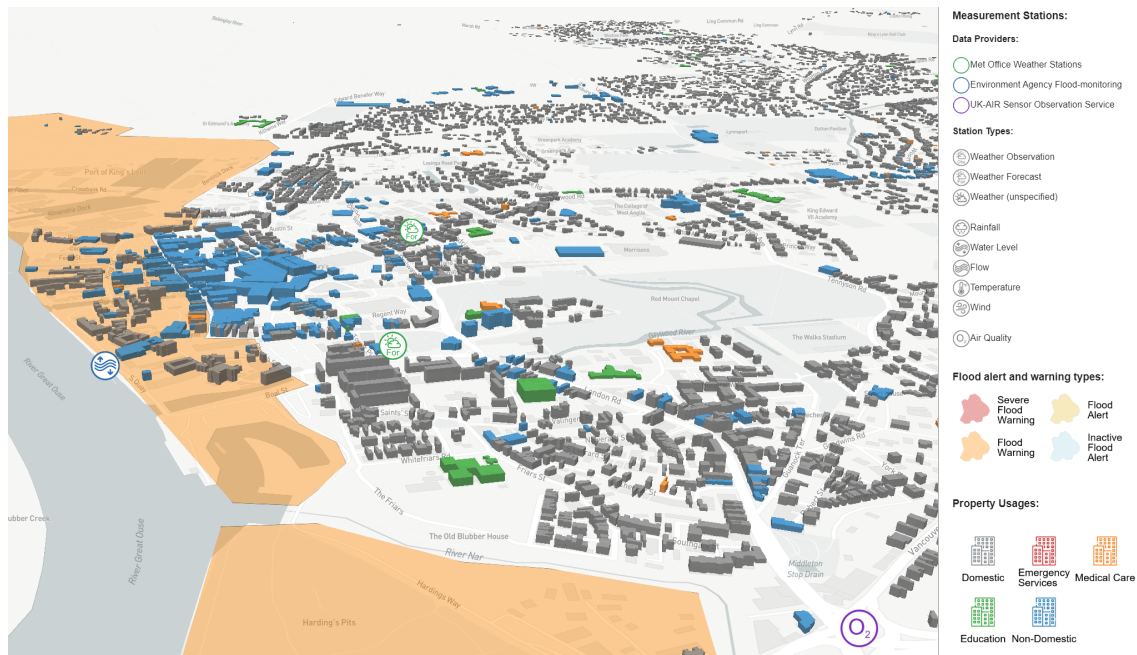


Figure 7: Building types at risk of flooding. The World Avatar provides insights about the usage types of potentially affected buildings to understand whether, *e.g.*, hospitals, emergency services or schools are affected. This information is crucial for tailored and timely decision-making and crisis response.

Besides this quantitative assessment, TWA provides further insights into the usage categories of potentially affected buildings to understand whether hospitals, emergency services or schools are at risk (depicted in Fig. 7). This information is crucial to tailor crisis responses and ensure timely decision-making. Furthermore, TWA provides the ability to store the flood warning history of flood areas in order to understand systematic risks and their trends, such as increasing or decreasing frequency or severity. Similar to the analysis of near misses in incident prevention, this information enhances understanding whether critical facilities, like hospitals or emergency services, are situated in locations that are increasingly prone to flooding or have previously been affected. Presently, this valuable information is currently unavailable, since all flood warnings provided by the real-time flood monitoring API are transient URIs (*i.e.*, once a flood warning is lifted, it simply gets deleted).

4.3 Extensible system of systems approach

So far, this work has demonstrated the automated assessment of immediate impacts of potential flooding events to facilitate short-term disaster response. However, this represents only one perspective to cope with flood hazards. Equally important is a thorough understanding of long-term risk exposure and the ability to conduct strategic planning and scenario analyses to improve flood preparedness and overall system resilience. The previously developed CReDo digital twin addresses exactly this gap: It uses simulated flood maps (*i.e.*, developed based on historical flooding events as well as climate change projections) to understand the impact of flooding on energy, water and telecoms network. Based on certain asset failure models, it provides insights into network interdependencies and investigates failure propagation both within and across networks. Since both works are based on TWA, they can easily be combined to provide a more comprehensive perspective to analyse flood risk and improve both strategic and tactical decision-making. This additional angle towards flood preparedness is illustrated in **Fig. 8**, where a severe flood warning is linked with an electrical substation located in the affected area to understand how a flood-imposed failure of this station would affect connected network assets. Such insights help refine potential hazard scenarios, such as whether to expect a failure of the telecoms network, and enable the early detection of potential weak points that require special attention during crisis response.

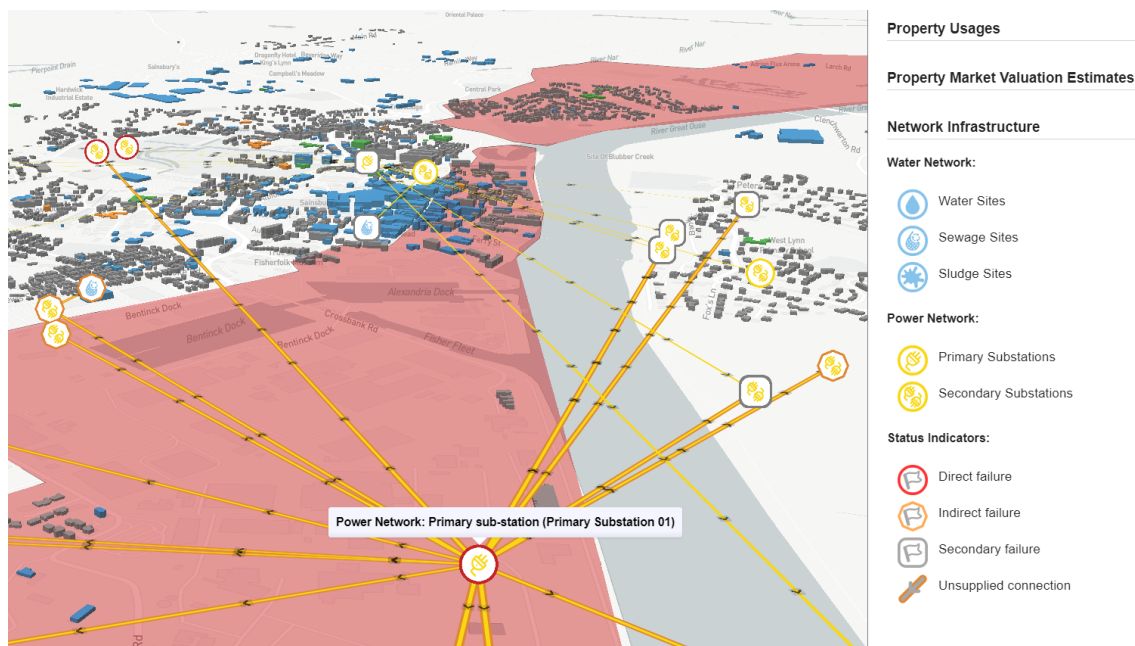


Figure 8: *Extensibility of the World Avatar. Based on knowledge models, TWA represents a suitable system to connect a variety of individual digital twins to combine cross-domain data and enable multi-perspective analyses. Beyond the dynamic assessment of potential impact of immediate flood risks, the very same system also enables strategic planning and scenario analyses to understand potential failure propagation of network infrastructure due to flooding: Illustrated for an electrical substation outage due to flooding, incl. all further failure cascading.*

The World Avatar is designed as extensible knowledge model based system to integrate and connect diverse digital twins within one unified ecosystem of data and agents. Based on this system of systems approach, it is able to represent complex systems without the need of having one huge monolithic model or platform. Instead, multiple connected digital twins provide different complementary perspectives on a certain domain or problem to enrich individual capabilities and analyses, such as combining tactical flood management with a more strategic flood preparedness dimension, or enriching rather static network failure analyses with further information about likely affected people and buildings in the vicinity of out-of-service assets.

5 Conclusions

This work demonstrates the benefits of knowledge model based approaches, such as the World Avatar dynamic knowledge graph, during disaster situations when timely and evidence-based cross-domain insights and decisions are required. An ecosystem of autonomous software agents continuously ingests and connects previously fragmented yet related real-world data from multiple APIs (*i.e.*, environmental sensor networks, energy performance certificate data, and property sales data) to enhance the value of individual pieces of information and derive more holistic perspectives on imminent flooding scenarios. TWA's unified visualisation framework provides aligned and consolidated access to these previously isolated data and empowers people with a more comprehensive view of the world and current situation around them, for instance a consolidated view of environmental observations, including weather, air quality, and river level readings. Additionally, the interplay of multiple interconnected agents provides an infrastructure to ensure automated information cascading, which is showcased for the automatic re-assessment of potential impacts of raised flood warnings with regards to people and buildings at risk whenever relevant input data changes. Although multiple flood monitoring tools, maps, and even warning services exist, this dynamism and cross-domain assessment represents a novelty to the best of our knowledge.

This showcase also highlights the potential of the World Avatar as a versatile flood management and impact assessment tool supporting both tactical and strategic decision-making. While most other tools focus solely on flood monitoring or strategic risk planning, the World Avatar is highly extensible and suited to provide multiple perspectives on complex cross-domain issues. While the system has initially been designed for evaluating the immediate impact of anticipated floods, it can also incorporate a long-term perspective by integrating strategic flood maps to analyse and improve climate adaptation and resilience. The Climate Resilience Demonstrator, previously developed by the UK National Digital Twin programme, can be integrated seamlessly to provide insights into infrastructure interdependencies and how potential flood-induced system failures could propagate both within but also beyond individual networks.

There are a few limitations of the current approach. Firstly, the system only focuses on flood alerts and warnings rather than evaluating actual flood events. Secondly, the prediction of flood events is not performed as part of the World Avatar, but relies on external warnings issued by domain experts from the Environmental Agency. Given the complex-

ity of hydrodynamic flood models, this is unlikely to change in the near future and is also not essential to demonstrate dynamic impact assessment. Thirdly, the current impact assessment is limited to the number of people and buildings at risk, as well as the total building stock value. Further assets such as roads or other infrastructure components are not yet included. Notably, the analysis only considers instantiated buildings, which may result in an underestimation of the total property value at risk. In addition, the implemented property market value estimation method could be considered rudimentary, although it is deemed adequate for this proof-of-concept stage. Lastly, the fuzzy address matching currently used to link properties with historical sales transactions can further be improved, particularly in light of the recent advancement in large language models.

Research data

All the codes developed are available on The World Avatar GitHub repository <https://github.com/cambridge-cares/TheWorldAvatar>.

Acknowledgements

This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. A part of this study has been undertaken in the context of DOME 4.0 project, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 953163. M. Hofmeister acknowledges financial support provided by the Cambridge Trust and CMCL. M. Kraft gratefully acknowledges the support of the Alexander von Humboldt Foundation.

Attribution statements:

This work contains OS data © Crown copyright and database rights 2022 (100025252). Furthermore, it contains HM Land Registry data © Crown copyright and database right 2020 as well as public sector information licensed under the Open Government Licence v3.0. Lastly, this showcase uses Environment Agency flood and river level data from the real-time data API (Beta).

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Nomenclature

- API** Application Programming Interface
- BHA** Building Height Attribute (OS Premium dataset)
- CReDo** Climate Resilience Demonstrator
- EA** Environment Agency

EPC Energy Performance Certificate
GeoSPARQL Geographic Query Language for RDF Data
GIS Geographic Information System
KG Knowledge Graph
OS Ordnance Survey
PM₁₀ Particulate Matter less than 10 μm in diameter
PM_{2.5} Particulate Matter less than 2.5 μm in diameter
SOS Sensor Observation Service
TWA The World Avatar
UK-AIR UK Air Information Resource
UKHPI UK House Price Index
UPRN Unique Property Reference Number
URI Uniform Resource Identifier

A Appendix

A.1 UK House Price Index

The UKHPI is produced by the Office for National Statistics (ONS) and represents a comprehensive housing market index that captures changes in house prices, market trends, and regional variations thereof across the UK. It utilises a hedonic regression model and mix-adjustment techniques to account for composition and property type changes between reporting periods. Data from various sources, such as Council Tax Valuation, EPCs, and Acorn classification, are incorporated to capture key construction, property, and demographic characteristics (details about the calculation methodology are provided by HM Land Registry [32]). Derived values represent nominal house price changes and are not inflation adjusted. The index covers all residential properties to provide an accurate and timely measure of the UK housing market, dating back to January 1995. Property post-codes are mapped to higher-level geographies using the ONS Postcode Lookup and Postcode Directory to calculate the index at national and regional level, as well as for counties, local authorities and London boroughs.

A.2 Population Data Comparison

To ensure meaningfulness of the slightly dated OpenPopGrid [40] population raster data, a comparison with the coarser, but continuously updated HDX [24] population map has been conducted on the Super Output Area Middle Layer. The relative error of just around 4% suggests that the OpenPopGrid still describes the current population distribution sufficiently accurate (see Fig. 9 and Table 3).

Table 3: Comparison of OpenPopGrid population raster with 2011 ONS census data as well as continuously updated HDX population map for three Super Output Area Middle Layer areas.

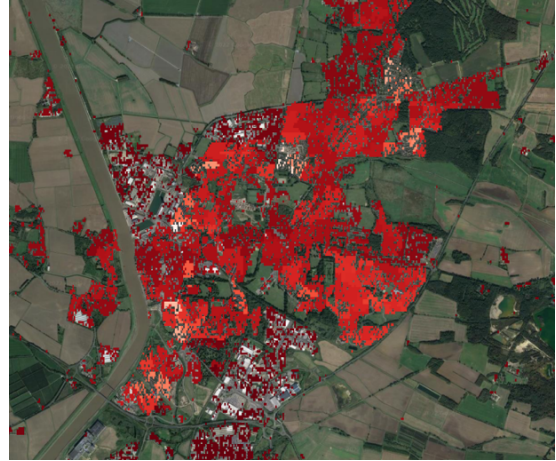
Super Output Area Middle Layer	Population estimates			Relative error	
	ONS Census 2011	OpenPopGrid	HDX population map	OpenPopGrid vs. HDX	OpenPopGrid vs. ONS
King's Lynn and West Norfolk 007	6701	5964	6287	-5.1%	-11.0%
King's Lynn and West Norfolk 009	7206	6946	7196	-3.5%	-3.6%
King's Lynn and West Norfolk 011	11142	9726	10074	-3.5%	-12.7%
			Average	-4.0%	-9.1%

A.3 UPRN Coverage

For each instantiated building, the OS Features API is queried with the building's bounding box to retrieve all enclosed UPRNs. Those are then tested against the building's



(a) OpenPopGrid data for King's Lynn centre



(b) HDX data for King's Lynn centre



(c) OpenPopGrid data for mix of residential and industrial area



(d) HDX data for mix of residential and industrial area

Figure 9: Visual comparison of OpenPopGrid and HDX population raster data for the town of King's Lynn: OpenPopGrid data has higher geospatial granularity and resembles actual residential areas more precisely.

GroundSurface and *lodFootPrint* geometries to further filter for intersecting UPRNs only. Remaining UPRNs are then instantiated and linked with the respective building(s). As illustrated in **Fig. 10**, not all buildings are associated with UPRN information. This could be the result of UPRNs falling within the bounds of a building but not intersecting its footprint, *e.g.*, those in the inner courtyards of a property; however, this cannot be confirmed from the figure. Hence, the sparseness of UPRN designations is likely attributed to limited UPRN coverage in available OS OpenData.



Figure 10: Comparison of buildings with (green) and without (red) available UPRN information: For buildings with missing UPRN, no identifier intersecting the building's footprint could be retrieved from OS Features API.

References

- [1] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal Digital Twin - A Dynamic Knowledge Graph. *Data-Centric Engineering*, 2, 2021. doi:10.1017/dce.2021.10.
- [2] J. Bai, K. F. Lee, M. Hofmeister, S. Mosbach, J. Akroyd, and M. Kraft. A derived information framework for a dynamic knowledge graph and its application to smart cities, 2023. Submitted for publication. Available from <https://como.ceb.cam.ac.uk/preprints/302/>.
- [3] P. D. Bates, N. Quinn, C. Sampson, A. Smith, O. Wing, J. Sosa, J. Savage, G. Olcese, J. Neal, G. Schumann, L. Giustarini, G. Coxon, J. R. Porter, M. F. Amodeo, Z. Chu, S. Lewis-Gruss, N. B. Freeman, T. Houser, M. Delgado, A. Hamidi, I. Bolliger, K. McCusker, K. Emanuel, C. M. Ferreira, A. Khalid, I. D. Haigh, A. Couasnon, R. Kopp, S. Hsiang, and W. F. Krajewski. Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, 57(2), Feb. 2021. doi:10.1029/2020wr028673.
- [4] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):28–37, 2001.
- [5] C. Bizer, T. Heath, and T. Berners-Lee. Linked data: The story so far. In A. Sheth, editor, *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. IGI Global, 2011. doi:10.4018/978-1-60960-593-3.ch008.
- [6] A. Chadzynski, S. Li, A. Grisiute, J. Chua, J. Yan, H. Y. Tai, E. Lloyd, M. Agarwal, M. Hofmeister, J. Akroyd, P. Herthogs, and M. Kraft. Semantic 3D City Interfaces - intelligent interactions on Dynamic Geospatial Knowledge Graphs, 2022. Submitted for publication. Available from <https://como.ceb.cam.ac.uk/preprints/297/>.
- [7] A. Chadzynski, S. Li, A. Grisiute, F. Farazi, C. Lindberg, S. Mosbach, P. Herthogs, and M. Kraft. Semantic 3D City Agents—An intelligent automation for dynamic geospatial knowledge graphs. *Energy and AI*, 8:100137, 2022. doi:10.1016/j.egyai.2022.100137.
- [8] A. Chadzynski, H. Silvennoinen, A. Grisiute, F. Farazi, S. Mosbach, M. Raubal, P. Herthogs, and M. Kraft. Semantic 3D City Inferences - multi-domain reasoning on Dynamic Geospatial Knowledge Graphs, 2023. Submitted for publication. Available from <https://como.ceb.cam.ac.uk/preprints/303/>.
- [9] H. Chang, A. Pallathadka, J. Sauer, N. B. Grimm, R. Zimmerman, C. Cheng, D. M. Iwaniec, Y. Kim, R. Lloyd, T. McPhearson, B. Rosenzweig, T. Troxler, C. Welty, R. Brenner, and P. Herreros-Cantis. Assessment of urban flood vulnerability using the social-ecological-technological systems framework in six US cities. *Sustainable Cities and Society*, 68:102786, 2021. doi:10.1016/j.scs.2021.102786.
- [10] Codestream Solutions Pte Ltd. FloodMapp, 2023. Available at <https://www.floodmapp.com/> (accessed May 2023).

- [11] Department for Environment Food & Rural Affairs. Real Time flood-monitoring API, 2021. Available at <https://environment.data.gov.uk/flood-monitoring/doc/reference> (accessed Feb 2023).
- [12] Department for Environment Food & Rural Affairs. Hydrology Data API, 2021. Available at <https://environment.data.gov.uk/hydrology/doc/reference> (accessed Apr 2023).
- [13] Department for Levelling Up, Housing & Communities. Energy Performance of Buildings Data, 2022. Available at <https://epc.opendatacommunities.org/docs/api> (accessed Nov 2022).
- [14] A. Devanand, G. Karmakar, N. Krdzavac, R. Rigo-Mariani, E. Y. S. Foo, I. A. Karimi, and M. Kraft. OntoPowSys: A power system ontology for cross domain interactions in an eco industrial park. *Energy and AI*, 1:100008, 2020. doi:10.1016/j.egyai.2020.100008.
- [15] Digital Twin Hub. Climate Resilience Demonstrator, 2022. Available at <https://digitaltwinhub.co.uk/credo/credo/> (accessed May 2023).
- [16] J. Dorman and N. Banerjee. A Real-Time Flood Warning System, 2016. Available at <https://www.esri.com/about/newsroom/wp-content/uploads/2018/09/a-real-time-flood-warning-system.pdf> (accessed May 2023).
- [17] EDINA Digimap Ordnance Survey Service, 2022. Available at <https://digimap.edina.ac.uk/> (accessed Nov 2022).
- [18] O. Ekmekcioğlu, K. Koc, and M. Özger. Towards flood risk mapping based on multi-tiered decision making in a densely urbanized metropolitan city of Istanbul. *Sustainable Cities and Society*, 80:103759, 2022. doi:10.1016/j.scs.2022.103759.
- [19] Environment Agency. Flooding in England: A National Assessment of Flood Risk, 2009. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/292928/geho0609bqds-e-e.pdf (accessed May 2023).
- [20] Esri Disaster Response Program. Flooding, 2023. Available at <https://esri-disasterresponse.hub.arcgis.com/pages/flooding> (accessed May 2023).
- [21] European Commission. European Union Floods Directive, 2023. Available at https://environment.ec.europa.eu/topics/water/floods_en (accessed May 2023).
- [22] European Commission, Copernicus. European Flood Awareness System, 2023. Available at <https://www.efas.eu/en> (accessed May 2023).
- [23] European Environment Information and Observation Network. General Multilingual Environmental Thesaurus, 2021. Available at <https://dd.eionet.europa.eu/vocabulary/aq/pollutant/view> (accessed Feb 2022).

- [24] Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. United Kingdom: High Resolution Population Density Maps + Demographic Estimates. High Resolution Settlement Layer (HRSL), 2016. Available at <https://data.humdata.org/dataset/united-kingdom-high-resolution-population-density-maps-demographic-estimates/resource/9007503c-5bf3-450f-8f3f-ca06682f0192> (accessed May 2023).
- [25] First Street Foundation. Flood Factor - Defining America's Past, Present, and Future Flood Risk, 2023. Available at <https://firststreet.org/risk-factor/flood-factor/> (accessed May 2023).
- [26] Flood Forecasting Centre. FGS Public API, 2017. Available at <https://api.foresources.metoffice.gov.uk/docs/flood-guidance-statement-api-public#> (accessed Apr 2023).
- [27] Flood Forecasting Centre. Flood Guidance Statement User Guide, 2017. Available at <https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/business/public-sector/hazard-manager/fgs-user-guide-january-2017-version-4.pdf> (accessed Apr 2023).
- [28] V. Gill and K. Stephens. Climate change: Warming could raise UK flood damage bill by 20%, 2023. Available at <https://www.bbc.com/news/science-environment-64866058> (accessed May 2023).
- [29] M. Hammond, A. Chen, S. Djordjević, D. Butler, and O. Mark. Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, 12(1):14–29, 2013. doi:10.1080/1573062x.2013.857421.
- [30] HM Land Registry. HM Land Registry Open Data, 2022. Available at <https://landregistry.data.gov.uk/> (accessed Nov 2022).
- [31] HM Land Registry. Price Paid Linked Data, 2022. Available at <https://landregistry.data.gov.uk/app/root/doc/ppd> (accessed Nov 2022).
- [32] HM Land Registry. UK House Price Index Linked Data, 2022. Available at <https://landregistry.data.gov.uk/app/ukhpi/doc> (accessed Nov 2022).
- [33] M. Hofmeister, J. Bai, G. Brownbridge, S. Mosbach, M. Hillman, F. Farazi, M. Agarwal, S. Ganguly, J. Akroyd, and M. Kraft. Semantic Agent Framework for Automated Flood Assessment Using Dynamic Knowledge Graphs, 2023. In preparation.
- [34] J. M. Johnson, T. Narock, J. Singh-Mohudpur, D. Fils, K. C. Clarke, S. Sak-sena, A. Shepherd, S. Arumugam, and L. Yeghiazarian. Knowledge graphs to support real-time flood impact evaluation. *AI Magazine*, 43(1):40–45, 2022. doi:10.1002/aaai.12035.
- [35] M. Kraft and S. Mosbach. The future of computational modelling in reaction engineering. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924):3633–3644, 2010. doi:10.1098/rsta.2010.0124.

- [36] Met Office. MetOffice DataPoint API, 2022. Available at <https://www.metoffice.gov.uk/services/data/datapoint/api-reference> (accessed Nov 2022).
- [37] Met Office and Environment Agency. Check for flooding, 2022. Available at <https://check-for-flooding.service.gov.uk/river-and-sea-levels> (accessed Apr 2023).
- [38] W. H. M. W. Mohtar, J. Abdullah, K. N. A. Maulud, and N. S. Muhammad. Urban flash flood index based on historical rainfall events. *Sustainable Cities and Society*, 56:102088, 2020. doi:10.1016/j.scs.2020.102088.
- [39] M. H. Mughal, Z. A. Shaikh, A. I. Wagan, Z. H. Khand, and S. Hassan. ORFFM: An ontology-based semantic model of river flow and flood mitigation. *IEEE Access*, 9:44003–44031, 2021. doi:10.1109/ACCESS.2021.3066255.
- [40] A. Murdock, A. Harfoot, D. Martin, S. Cockings, and C. Hill. OpenPopGrid: an open gridded population dataset for England and Wales. GeoData, University of Southampton, 2015. Available at <http://openpopgrid.geodata.soton.ac.uk/> (accessed Apr 2023).
- [41] M. R. Najafi, Y. Zhang, and N. Martyn. A flood risk assessment framework for interdependent infrastructure systems in coastal environments. *Sustainable Cities and Society*, 64:102516, 2021. doi:10.1016/j.scs.2020.102516.
- [42] Ordnance Survey. Ordnance Survey Data Hub, 2022. Available at <https://osdatahub.os.uk/> (accessed Nov 2022).
- [43] S. H. Pour, A. K. A. Wahab, S. Shahid, M. Asaduzzaman, and A. Dewan. Low impact development techniques to mitigate the impacts of climate-change-induced urban floods: Current trends, issues and challenges. *Sustainable Cities and Society*, 62:102373, 2020. doi:10.1016/j.scs.2020.102373.
- [44] A. Rebally, C. Valeo, J. He, and S. Saidi. Flood impact assessments on transportation networks: A review of methods and associated temporal and spatial scales. *Frontiers in Sustainable Cities*, 3, 2021. doi:10.3389/frsc.2021.732181.
- [45] T. Savage, J. Akroyd, S. Mosbach, M. Hillman, F. Sielker, and M. Kraft. Universal digital twin – the impact of heat pumps on social inequality. *Advances in Applied Energy*, 5:100079, 2022. doi:10.1016/j.adapen.2021.100079.
- [46] T. Savage, J. Akroyd, S. Mosbach, N. Krdzavac, M. Hillman, and M. Kraft. Universal digital twin: Integration of national-scale energy systems and climate data. *Data-Centric Engineering*, 3, 2022. doi:10.1017/dce.2022.22.
- [47] T. Sieg and A. H. Thielen. Improving flood impact estimations. *Environmental Research Letters*, 17(6):064007, 2022. doi:10.1088/1748-9326/ac6d6c.
- [48] UK-AIR Sensor Observation Service (Beta release). Department for Environment Food & Rural Affairs, 2023. Available at https://uk-air.defra.gov.uk/data/about_sos (accessed Apr 2023).

- [49] United Nations. United nations sustainable development goals, 2021. Available at <https://sdgs.un.org/goals> (accessed May 2023).
- [50] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, and et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1): 160018, 2016. doi:10.1038/sdata.2016.18.
- [51] D. D. Wrachien, J. Garrido, S. Mambretti, and I. Requena. Ontology For Flood Management: A Proposal. In D. Proverbs, S. Mambretti, C. Brebbia, and D. D. Wrachien, editors, *Flood Recovery, Innovation and Response III*. WIT Press, 2012. doi:10.2495/FRIAR120011.
- [52] X. Zhou, A. Eibeck, M. Q. Lim, N. Krdzavac, and M. Kraft. An agent composition framework for the J-Park Simulator – a knowledge graph for the process industry. *Comput. Chem. Eng.*, 130:106577, 2019. doi:10.1016/j.compchemeng.2019.106577.