Mapping the Built Environment Process (BEP) Ecosystem via a Data to Knowledge Framework

NAOMI KEENA
Yale Center for Ecosystems in Architecture, Yale University

ANNA DYSON
Yale Center for Ecosystems in Architecture, Yale University

MOHAMED ALY ETMAN
Yale Center for Ecosystems in Architecture, Yale University

Keywords: Built Environment Process, Visual Analytics, Big Data, Data Visualization, Knowledge Frameworks, Socio-Ecological Design

Transitioning to a future of low-carbon built environments requires the design of multi-beneficial design strategies that take a whole building life cycle and systems-thinking approach. Such an approach has the potential to enable multi-stakeholder engagement and cross-industry collaboration which are current siloed in the Built Environment Process (BEP). The BEP involves energy, material and information flows at each of its phases from the initial extraction of raw materials to the final deconstruction of a building. Technology and big data have a role to play in establishing collaborative networks with efficient construction practices which track material, energy and information flows across the building life cycle. This paper attempts to map the BEP through a new data to knowledge framework named SEVA (Socio-Ecological Visual Analytics), which has been designed to link heterogeneous data. It describes the methodology used to map the BEP in SEVA. This involves the deployment of semantic web ontologies to generate a knowledge graph of the BEP; virtually connecting each phase and its associated stakeholders, thereby, conceivably acting as an overview tool for the BEP. As climate pressures increase and material scarcity is imminent, innovation in eco-systems thinking and data to knowledge frameworks will be critical towards ensuring built environments embrace a socio-ecological future.

SCOPING THE BUILT ENVIRONMENT PROCESS (BEP): ENVIRONMENTAL IMPACTS AND ORGANIZATIONAL BARRIERS WITHIN THE BEP

According to federal scientists, understanding the consequences of climate change on the US involves studying the interconnections between the natural, built, and social systems we rely on and their vulnerability to cascading impacts (USGCRP 2018) . Notwithstanding this complexity, as pointed out by AIA's "Designing for Integration" measure (AIA 2020), individual design strategies can offer multi-faceted value across social, economic, and environmental systems. Managing interconnections between systems poses many challenges, including linking siloed streams of heterogeneous data, uniting various stakeholders, and necessitating intellectual agility to respond to societal, economic, and environmental shifts.

This paper outlines ongoing interdisciplinary research, exploring the harnessing of big data in mapping interconnections within the BEP (Keena and Dyson 2017; 2020; Keena, 2017). By tracking carbon, energy and material flows, it aims to surpass the concept of a building, in abstraction, fixed solely in the operational phase, but rather as a system which undergoes multiple journeys of carbon, energy and material transformation in its initial construction and future dismantle. Such a system includes many stakeholders who represent each phase of the BEP. According to the Department of Energy (DOE 2008) , the compartmentalization and lack of communication between building professionals in each sector results in suboptimal designs and less than optimal building operations while contributing to environmental impacts (USHUD 2003; Du Plessis and Cole 2011). A McKinsey report on the construction sector echoes this view, defining the sector's lack of productivity and predicting that, faced with sustainability demands, the sector will need to reassess how it builds to reduce waste and abate carbon emissions (Barbosa, Woetzel, and Mischke 2017) .The report also highlights the role big data can play. In the construction phase alone, the report predicts that an increase of up to 50 percent on-site productivity could be attributed to the implementation of data techniques and accurate data flows through various stakeholder systems that are both *backward looking and predictive*. Within the entire BEP, data to knowledge frameworks have a significant role to play towards an ecosystem intelligence enabling distributed knowledge across life cycle phases. This is particularly relevant to overcome the challenge of the expanded scope of the BEP with distributed teams and complex information flows, as illustrated in Figure 1.

METHODOLOGY: A DATA TO KNOWLEDGE FRAMEWORK

Bridging the gap between building stakeholders and navigating a multi-scalar expanded scope of design may have been unforeseen in the 20th Century, but with a transition from industrial societies to knowledge societies, today data to knowledge frameworks offer unprecedented opportunities in decoding complexity (UN Environment 2019). The

Figure 1. Towards inclusive environments that link data, design and knowledge into parametric, queryable environments that can constantly be reconfigured according to new information. Image courtesy of authors.

methodology employed includes mapping the BEP through a new data to knowledge framework named SEVA (Socio-Ecological Visual Analytics) (Aly Etman, Keena and Dyson 2020; Keena and Dyson, 2017; Keena 2017; Aly Etman, 2018), which has been designed to link heterogeneous data. Using web ontology language (McGuinness and Van Harmelen 2004) and semantic web frameworks, a knowledge graph of the BEP is generated in SEVA. The concept of a knowledge graph involves creating meaningful virtual connections and relationships between data, representing each BEP phase and its associated stakeholders.

These *virtual connections* mapping the BEP are made through semantic web frameworks (Berners-Lee, Hendler and Lassila 2001). SEVA employs two open-source semantic web frameworks: 1) Whyis, and 2) Human-Aware Data Acquisition (HADatAc), as illustrated in Figure 2. Whyis is a framework to publish, manage and analyze nanopublications inside of a knowledge graph (McCusker et al. 2018; 2020). In this case nanopublication is defined as the smallest unit of publishable information, an assertion about any topic that can be uniquely identified and attributed to its author. HADatAc, a data and metadata repository, is used to integrate, overlay, and link metadata to raw data and scientific annotation (Pinheiro et al. 2018). HADatAc allows for capturing knowledge about how acquired data was generated, measured or informed. HADatAc links the data to the metadata.

Both Whyis and HADatAc are enabled by ontologies. Ontologies, described in greater detail elsewhere (Gruber 1993; McGuinness and Van Harmelen 2004; Bizer, Heath, and Berners-Lee, 2011), can be defined as the explicit formal specification of the terms in a domain and relations among them; they provide a way to encode meaning that computers and people can unambiguously understand. The SEVA environment generates connections across data at all levels with the help of its own built environment ontology named 'Built Ecologies Ontology' (BEO) which is a family of existing ontologies including: (1) Human-Aware Science Ontologies (HAScO) which integrates a collection of well-established science-related ontologies and aims to address issues related to data annotation for a large data ecosystem, where data can come from diverse data sources from sensors to questionnaires, (Pinheiro et al. 2018; Santos et al. 2017); (2) Semanticscience Integrated Ontology (SIO) provides a simple, integrated ontology of types and relations for rich description of objects, processes and their attributes (Dumontier et al. 2014). (3) PROV-O represents provenance information for different applications, domains and under difference contexts, (Lebo et al. 2013). (4) Virtual Solar Terrestrial Observatory (VSTO), represents observational, experimental, and model databases in the fields of solar, solar-terrestrial, and space physics, (Fox et al. 2009). (5) Environment Ontology (ENVO), represents the concise, controlled description of environments (Buttigieg, P. L., et al 2013; 2016). (6) United Nations Sustainable Development Goals Interface Ontology (SDGIO), which aims to provide a semantic bridge between a) the Sustainable Development Goals, their targets, and indicators and b) the large array of entities they refer to (Smith and Jensen 2016; Jensen 2016; Buttigieg et al., 2016). (7) Units of measurement ontology (UO), represents information on various units of measurement (Gkoutos, Schofield and Hoehndorf 2012).

Each data point uploaded to SEVA is accompanied by relevant contextual knowledge related to that data point. This is then linked via semantic web frameworks (i.e. Whyis and HADatAc) to appropriate ontologies that classify and give meaning to that data point, including the ability to automatically generate and discover new relationships based not only on that data point, but also on the contextual knowledge and vocabulary about the data provided by data dictionaries; in other words, it allows the computer to logically process knowledge based on descriptive logic, i.e. inferences can be made upon it.

By linking BEP data in this way, a stakeholder can nimbly generate visualizations and data-storytelling by pulling data from an *intelligent* knowledge graph that has multiple connections and relationships already embedded within it. This aims to tackle

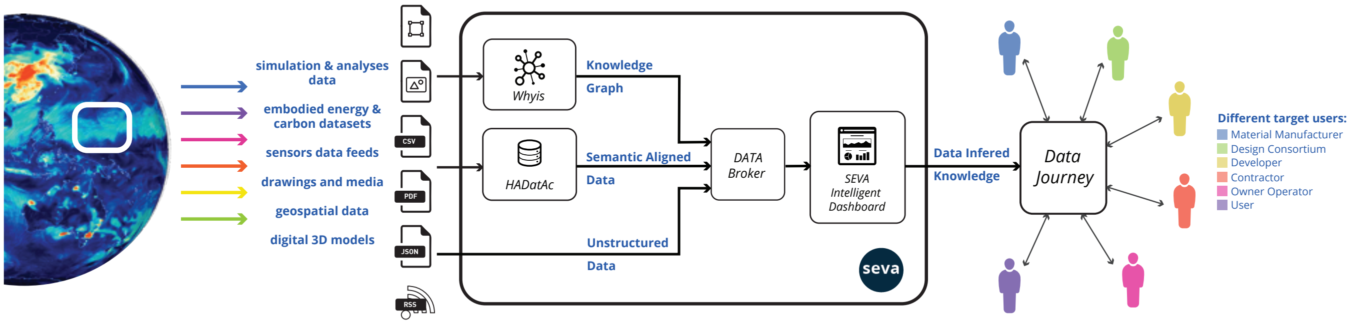

Figure 2. The SEVA knowledge framework: the alignment of meta-data and micro-data, the semantic overlaying and critical annotation for data-knowledge management and the visual analytics approach for data access, representation and data-storytelling. Image courtesy of authors

a specific challenge with big data, especially within the BEP, in that different levels of data (e.g. simulation and analyses data, embodied energy and carbon datasets, sensor-data, drawings and media, geospatial data, digital 3D models, sketches etc.) involve disparate heterogeneous data formats that are often shown together but not necessarily *linked* via a knowledge framework. SEVA aims to address this challenge by intelligently linking various types of heterogeneous data as described above, but it also strives to enable visualization of the linked data and the generation of *data journeys* for data storytelling and sharing, as shown in Figure 2. Through the creation of data stories at each phase and timeframe of the BEP, SEVA aims to provide visual clarity to the complexities of a multi-scalar BEP while mobilizing various stakeholder engagement.

POTENTIAL SCENARIOS: LINKING EACH PHASE OF THE BEP THROUGH VISUALIZED DATA JOURNEYS

Once these virtual connections are made and the data is ingested into SEVA – the linked-data can be visualized. Here, each phase of the BEP is demonstrated and visualized through SEVA to indicate potential scenarios of how the framework can help to organize and link these phases. These BEP phases are described below and illustrated in Figure 3.

The work of the Geo-biosphere, Material Sourcing and Manufacturing Phase: Using SEVA to map and visualize this phase involves tracking the material life cycle from the raw material sourcing location to the extraction and manufacturing processes. With each data journey, not only the relevant energy and material data is captured and visualized but also imagery of sourcing conditions, geospatial maps etc. in order to create transparency and empathy regarding material sourcing and its relationship to the geo-biosphere. Such mapping of information flows can help link the architectural work back to the original sources of raw materials from which it is created. It can help us understand and visualize the environmental impacts associated with a building project at the global scale.

The geo-biosphere is the global ecological system that sustains life on earth. Earth's geo-biosphere includes the atmosphere (air), lithosphere (rock), hydrosphere (water), biosphere (living organisms) (Odum 2002; Yang 2018). The acquisition of raw materials typically involves mining and quarrying materials (construction minerals, sand, gravel, crushed stone, cement) which are used in the production of common construction materials (e.g. steel, aluminum, concrete, glass). These are typically non-renewable resources which rely on long term natural ecological products and services, such as natural sedimentary cycles for their formation. These cycles can take anywhere from thousands to millions of years. This knowledge of the environmental context is represented via the ENVO and VTSO ontologies in particular and the potential sustainable development implications are represented via the SDGIO.

Our reliance on the work of the geo-biosphere is not only for raw materials but also for other ecosystem services (MEA 2005) such as the dilution of air pollution often associated with the manufacturing and production stages of construction materials. Mapping and tracking these interconnected information flows of the BEP ecosystem highlights the potential for positive (or mitigating negative) reinforcing feedback loops.

The Design Phase: By mapping the design phase in SEVA, an archive of design decisions and analysis is created. This bank of knowledge also facilitates project management via an online platform by integrating schematic design options, site and climate analyses, performance analyses (energy, life cycle, daylighting, structural etc.), 3D models, construction drawing-sets, renderings, animated walk-throughs etc. This offers transparency to the design team and becomes a tool where design members can share data narratives or journeys with their clients and specific stakeholders during the design phase.

Construction and Project Management Phase: The construction phase involves multiple stakeholders where project management can be challenging. The phase is known for its notoriously slow pace and a lack of productivity. Analyzing these limitations, a McKinsey report (Blanco, et al. 2018) explains that unlike other sectors, such as manufacturing and transportation which now operate more as ecosystems, the

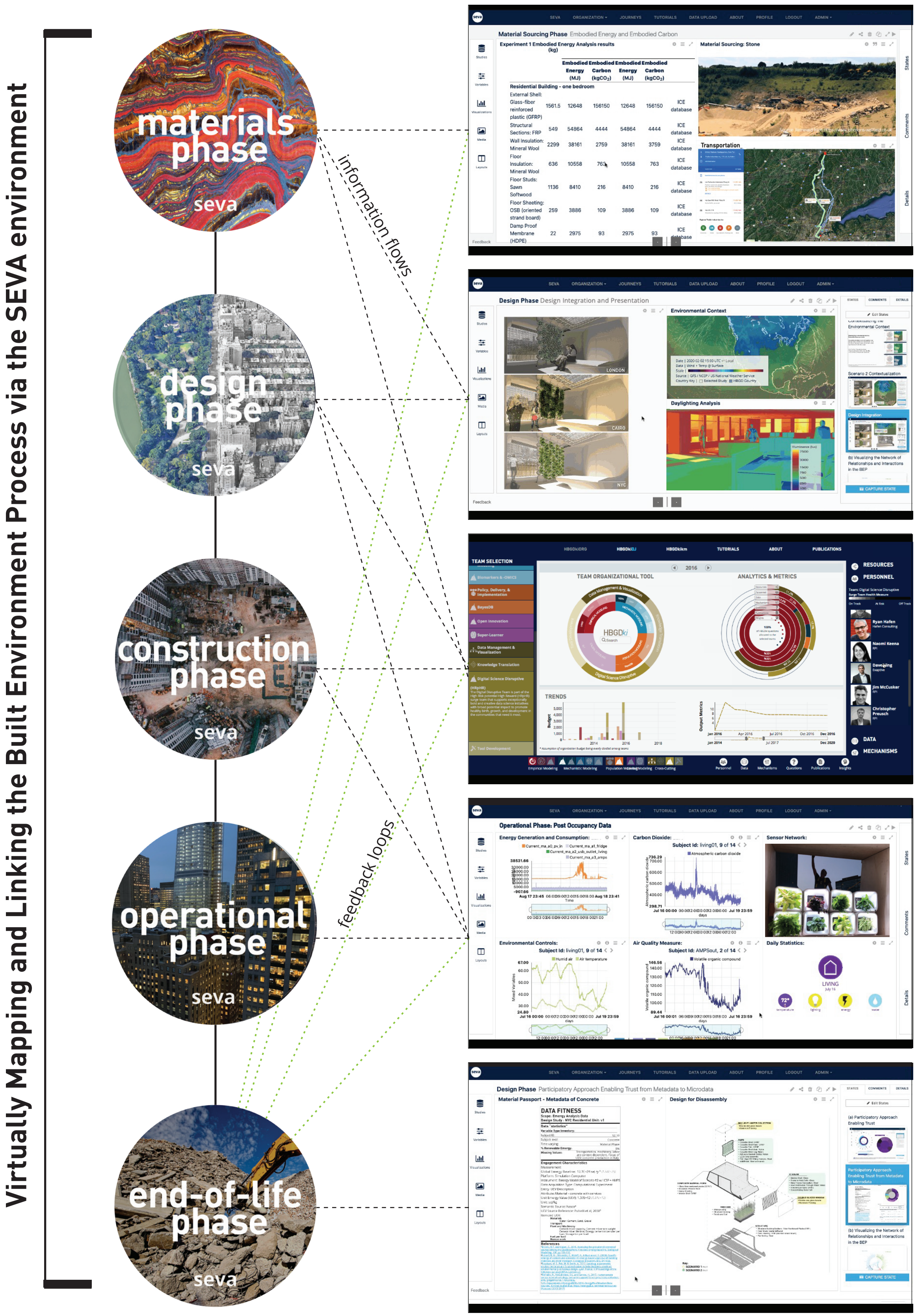

Figure 3. Each BEP phase is illustrated here as a snap-shot from visualized data journeys or illustrated narratives in SEVA. Each phase provides information flows that can form feedback loops at later phases in the whole building life cycle. Image courtesy of authors

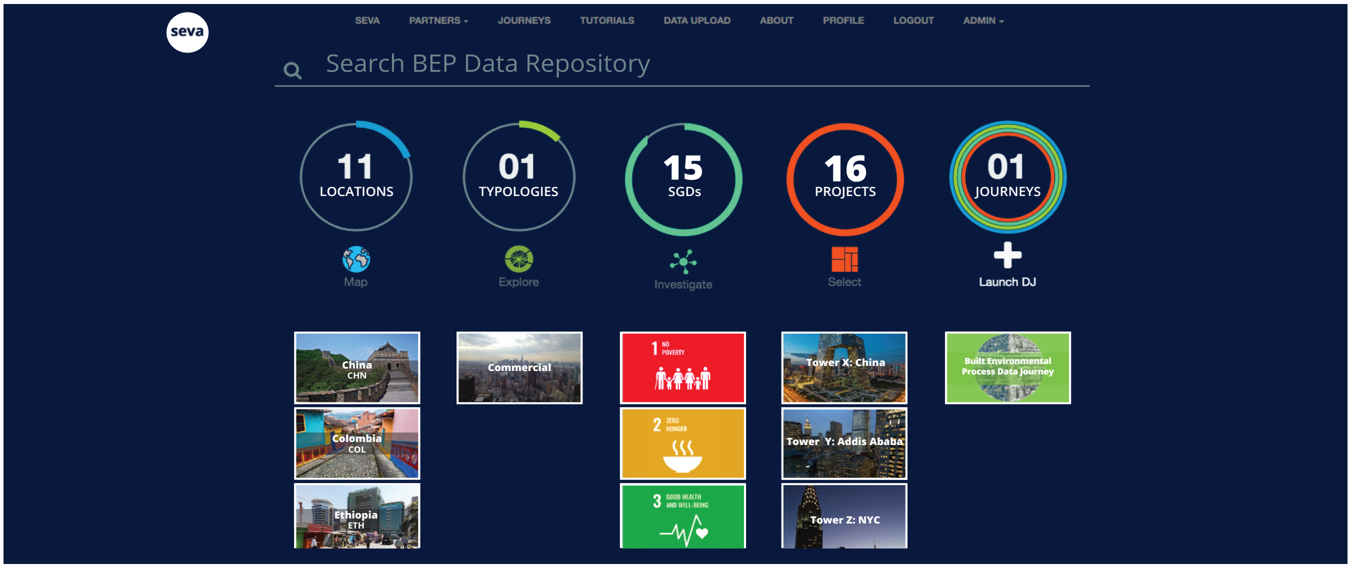

Figure 4. Visualizing and querrying the BEP data repository with a range of building projects filtered by a variety of factors.Image courtesy of authors

building sector continues to operate within siloes. It claims the sector's failure to embrace digital techniques have kept it "stuck in a time warp". It predicts that the sector will need to reassess such rigid organizational structures and adopt digitalization. The mapping at this stage within SEVA focuses on organization of construction teams, industry stakeholders, budget, schedules, digital drawings and models, as illustrated in Figure 3. The web-based platform facilitates the upload of live footage of construction sites towards on-site safety, site management, waste reduction and carbon abatement towards a more holistic project organizational system.

Operational and Maintenance Phase: The mapping of the operational and maintenance phase focuses on the collection of data via sensor networks and internet of things (IoT) to capture and visualize operational factors such as environmental conditions within the building, air quality monitoring and building system performance. This data once collected also catalogues post-occupancy data. Mapping this phase can enable feedback loops towards enhanced occupancy comfort.

End-of-Life Phase: In terms of mobilizing a circular economy in the built environment, knowledge from the earlier phases (as outlined above) is critical towards enabling the elimination of waste. For example, in order to re-use a material, knowledge and information about that material is crucial, such as how that material was initially sourced, how it was used in construction and maintained during the life cycle. Such knowledge facilitates informed decision making and is captured through the use of material passports (Heinrich and Lang 2019) . SEVA provides a 'Data Fitness Ticket' which adds a layer of transparency to the material passport by clearly identify sources of the data and providing links back to the different phases to elaborate on how this data was acquired – it helps characterize the uncertainty of the data. Additionally, any design for disassembly considerations during the development of construction assemblies, during the design phase, offers valuable knowledge for informed information during the deconstruction stage at end-of-life.

RESULTS: VISUAL MAPPING AND SEMANTIC LINKAGES TOWARDS A SEAMLESS BEP ECOSYSTEM

Alongside highlighting gaps, redundancy and environmental impacts within the BEP, the results show cross-cutting opportunities within BEP activities. Phases which are operating within siloes are highlighted incentivizing stakeholder mobilization towards breaking down such barriers and facilitating stakeholders of the building sector to operate more as an ecosystem across life cycle phases. Each phase has multiple data journeys or data narratives associated with them. Screen captures of states from these data journeys are captured in Figure 3 showing the extended life span of a building. Each data journey is embedded with information flows and potential feedback loops. Especially at the later stages, data mapped in the early design, construction and manufacturing stages has much influence and potential to enhance informed decision-making and possibilities at the end-of-life stage. In this way the data becomes a living document of the memory of a project. Figure 4 illustrates how such a process offers the potential of a repository and linked database of projects by location, typology and analysis, phases or building framework, that can then be easily queried offering a bank of knowledge on different architectural projects. In Figure 4, the data is analyzed in terms of the United Nations Sustainable Development Goals (SDGs) established framework.

SIGNIFICANCE

This research has significance in multi-stakeholder engagement and evidence-based decision-making, especially within work which strives to find solutions to grand challenges such as environmental issues. By unlocking the potential of big data for the BEP, it aims to facilitate in projecting future scenarios towards a sustainable and progressive future. It offers potential value to circular economy methods, amendments to policy and building codes, and the creation of incentives for cross industry collaboration.

ACKNOWLEDGEMENTS

This work was made possible by the kind support of UN Environmental through World Environment Situation Room, UN Environment and UN Habitat through Ecological Living Module and the Bill and Melinda Gates Foundation through HBGDki.

ENDNOTES

1. USGCRP, "Fourth National Climate Assessment: Impacts, Risks, and Adaptation in the United States." Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart eds. Washington, DC: U.S. Global Change Research Program, 2018, 1515 pp. doi: 10.7930/NCA4.2018.
2. AIA, "Designing for Integration" from "The Framework for Design Excellence". The American Institute of Architects (AIA), 2020. https://www.aia.org/resources/6077668-framework-for-design-excellence
3. Keena, Naomi, and Anna Dyson. "Qualifying the quantitative in the construction of built ecologies". In Embodied energy and design edited by D. Benjamin, 196-205. New York: Columbia University Press, GSAPP Lars Müller, 2017.
4. Keena, Naomi, and Anna Dyson. "State of play for circular built environment in North America: A report compiling the regional state of play for circularity in the built environment in North America across the United States of America". Yale Center for Ecosystems in Architecture, Yale University and United Nations One Planet Network Sustainable Buildings and Construction Programme, 2020.
5. Keena, Naomi."Towards a Comprehensive Visual Analytics Environment for the Assessment of Socio Ecological Factors within Architectural Design". Doctoral dissertation. Renssealear Polytechnic Institute, 2017
6. DOE. "Building Technologies Program, Planned Program Activities for 2008-2012". United States: Department of Energy, 2008.
7. USHUD. "Building Better Homes: Government Strategies for Promoting Innovation in Housing, U.S. Department of Housing and Urban Development, Office of Policy Development and Research and the Partnership for Advancing Technology in Housing". Prepared by Rand Corp, 2003.
8. Du Plessis, Chrisna, and Raymond J. Cole. "Motivating change: shifting the paradigm." Building Research & Information 39, no. 5, 2011: 436-449.
9. Barbosa, Filipe, Jonathan Woetzel, and Jan Mischke. "Reinventing Construction: A Route of Higher Productivity". McKinsey Global Institute, 2017.
10. UN Environment. "Building a Digital Ecosystem for the Planet-Foresight Brief No. 014 September 2019." Campbell, Jillian, and David Jensen. Foresight Briefs, 2019.
11. Aly Etman, Mohamed, Naomi Keena and Anna Dyson. "Socio-Ecological Visual Analytics Environment "SEVA": A novel visual analytics environment for interdisciplinary decision-making linking human biometrics and environmental data." In Proceedings of the next WSBE | World Sustainable Built Environment conference, BEYOND 2020. Gothenburg, Sweden, online conference, 2020.
12. Aly Etman, Mohamed, Naomi Keena and Anna Dyson. "SEVA - Socio-Ecological Visual Analytics Environment: A new visual analytics environment for interdisciplinary decision-making linking human biometrics and environmental data."Doctoral dissertation. Renssealear Polytechnic Institute, 2018
13. McGuinness, Deborah L., and Frank Van Harmelen. 2004. OWL web ontology language overview. W3C recommendation 10.10, 2004.
14. Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." Scientific american 284, no. 5, 2001: 34-43.
15. McCusker, James P., Sabbir M. Rashid, Nkechinyere Agu, Kristin P. Bennett, and Deborah L. McGuinness. "The Whyis Knowledge Graph Framework in Action." In International Semantic Web Conference (P&D/Industry/BlueSky). 2018.
16. McCusker, James P., Neha Keshan, Sabbir Rashid, Michael Deagen, Cate Brinson, and Deborah L. McGuinness. "NanoMine: A Knowledge Graph for Nanocomposite Materials Science." In International Semantic Web Conference, pp. 144-159. Springer, Cham, 2020.
17. Pinheiro, Paulo, Henrique Santos, Zhicheng Liang, Yue Liu, Sabbir M. Rashid, Deborah L. McGuinness, and Marcello Peixoto Bax. "HADatAc: A Framework for Scientific Data Integration using Ontologies." In International Semantic Web Conference (P&D/Industry/BlueSky). 2018.
18. Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked data: The story so far." In Semantic services, interoperability and web applications: emerging concepts, pp. 205-227. IGI Global, 2011.
19. Gruber, Thomas R. "A translation approach to portable ontology specifications." Knowledge acquisition 5, no. 2, 1993: 199-220.
20. Pinheiro, Paulo, Marcello Peixoto Bax, Henrique Santos, Sabbir M. Rashid, Zhicheng Liang, Yue Liu, Yarden Ne'eman, James P. McCusker, and Deborah L. McGuinness. "Annotating Diverse Scientific Data With HAScO." In ONTOBRAS, pp. 80-91. 2018.
21. Santos, Henrique, Victor Dantas, Vasco Furtado, Paulo Pinheiro, and Deborah L. McGuinness. "From data to city indicators: A knowledge graph for supporting automatic generation of dashboards." In European Semantic Web Conference, pp. 94-108. Springer, Cham, 2017.
22. Michel, Christopher JO Baker, Joachim Baran, Alison Callahan, Leonid Chepelev, José Cruz-Toledo, Nicholas R. Del Rio et al. "The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery." Journal of biomedical semantics 5, no. 1, 2014: 1-11.
23. Lebo, Timothy, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. "Prov-o: The prov ontology." W3C recommendation 30, 2013.
24. Fox, Peter, Deborah L. McGuinness, Luca Cinquini, Patrick West, Jose Garcia, James L. Benedict, and Don Middleton. "Ontology-supported scientific data frameworks: The virtual solar-terrestrial observatory experience." Computers & Geosciences 35, no. 4 (2009): 724-738.
25. Buttigieg, P., N. Morrison, B. Smith, C. Mungall, and S. Lewis. "The environment ontology: contextualising biological and biomedical entities." Journal of Biomedical Semantics, 4(1), 2013: 43. doi:10.1186/2041-1480-4-43;
26. Buttigieg, Pier Luigi, Evangelos Pafilis, Suzanna E. Lewis, Mark P. Schildhauer, Ramona L. Walls, and Christopher J. Mungall. "The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation." Journal of Biomedical Semantics, 7(1), 2016: 57. doi:10.1186/s13326-016-0097-6)
27. Smith, Barry and Mark Jensen, "The UNEP Ontologies and the OBO Foundry." ICBO/BioCreative, 2016.
28. Jensen, Mark. "Sustainable Development Goals Interface Ontology." In ICBO/BioCreative. 2016;
29. Buttigieg, Pier Luigi, Mark Jensen, Ramona L. Walls, and Christopher J. Mungall. "Environmental Semantics for Sustainable Development in an Interconnected Biosphere." In ICBO/BioCreative. 2016.
30. Gkoutos, Georgios V., Paul N. Schofield, and Robert Hoehndorf. "The Units Ontology: a tool for integrating units of measurement in science." Database 2012, 2012.
31. Odum, Howard T. "Material circulation, energy hierarchy, and building construction." Construction ecology: nature as the basis for green buildings, 2002: 37-71.
32. Yang, Qing, Gengyuan Liu, Marco Casazza, Elliot T. Campbell, Biagio F. Giannetti, and Mark T. Brown. "Development of a new framework for non-monetary accounting on ecosystem services valuation." Ecosystem services 34, 2018: 37-54.
33. MEA, Millennium Ecosystem Assessment. "Ecosystems and human well-being: synthesis." Island, Washington, DC, 2005.
34. Blanco, Jose Luis, Steffen Fuchs, Matthew Parsons, and Maria João Ribeirinho. "Artificial intelligence: Construction technology's next frontier." Building Economist, The Sep 2018, 2018: 7.
35. Heinrich, M., and W. Lang. "Materials Passports: Best Practice Innovative Solutions for a Transition to a Circular Economy in the Built Environment." Technische Universität München, Munich, 2019.