




Article

People, Projects, Organizations, and Products: Designing a Knowledge Graph to Support Multi-Stakeholder Environmental Planning and Design

Sean N. Gordon ^{1,*}, Philip J. Murphy ², John A. Gallo ³ , Patrick Huber ⁴ , Allan Hollander ⁴, Ann Edwards ⁵ and Piotr Jankowski ^{6,7} 

- ¹ Institute for Natural Resources, Oregon State University, Portland, OR 97207, USA
² InfoHarvest, Inc., Seattle, WA 98165, USA; philip.murphy@infoharvest.com
³ Conservation Biology Institute, Corvallis, OR 97333, USA; john.gallo@consbio.org
⁴ Department of Environmental Science and Policy, University of California Davis, Davis, CA 95616, USA; prhuber@ucdavis.edu (P.H.); adhollander@ucdavis.edu (A.H.)
⁵ Resilient Earth, Vashon, WA 98070, USA; ann@resilient-earth.com
⁶ Department of Geography, San Diego State University, San Diego, CA 92182, USA; pjankows@sdsu.edu
⁷ Institute of Geocology and Geoinformation, Adam Mickiewicz University, 61-680 Poznan, Poland
* Correspondence: sean.gordon@oregonstate.edu

Abstract: As the need for more broad-scale solutions to environmental problems is increasingly recognized, traditional hierarchical, government-led models of coordination are being supplemented by or transformed into more collaborative inter-organizational networks (i.e., collaboratives, coalitions, partnerships). As diffuse networks, such regional environmental planning and design (REPD) efforts often face challenges in sharing and using spatial and other types of information. Recent advances in semantic knowledge management technologies, such as knowledge graphs, have the potential to address these challenges. In this paper, we first describe the information needs of three multi-stakeholder REPD initiatives in the western USA using a list of 80 need-to-know questions and concerns. The top needs expressed were for help in tracking the participants, institutions, and information products relevant to the REDP’s focus. To address these needs, we developed a prototype knowledge graph based on RDF and GeoSPARQL standards. This semantic approach provided a more flexible data structure than traditional relational databases and also functionality to query information across different providers; however, the lack of semantic data expertise, the complexity of existing software solutions, and limited online hosting options are significant barriers to adoption. These same barriers are more acute for geospatial data, which also faces the added challenge of maintaining and synchronizing both semantic and traditional geospatial datastores.

Keywords: decision support software; open systems; human–environment systems; environmental planning; information needs assessment; ontology; knowledge graph



Citation: Gordon, S.N.; Murphy, P.J.; Gallo, J.A.; Huber, P.; Hollander, A.; Edwards, A.; Jankowski, P. People, Projects, Organizations, and Products: Designing a Knowledge Graph to Support Multi-Stakeholder Environmental Planning and Design. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 823. <https://doi.org/10.3390/ijgi10120823>

Academic Editor: Wolfgang Kainz

Received: 27 October 2021

Accepted: 27 November 2021

Published: 6 December 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Authority for managing environmental issues in the USA is distributed across local, state, and federal agencies, requiring “vertical” collaboration across multiple organizational levels, and environmental problems often cross jurisdictional boundaries, requiring “horizontal” coordination across these boundaries as well [1]. As the need for more broad-scale solutions to environmental problems is increasingly recognized, traditional hierarchical, government-led models of coordination are being supplemented by or transformed into more collaborative inter-organizational networks (i.e., collaboratives, coalitions, partnerships), which bring together multiple existing organizations using various coordinating mechanisms [2–5]. In a pioneering book on landscape ecology, Forman [6] promoted such regional solutions, where a region is an area composed of landscapes with the same macroclimate and tied together by human activities. The concept links the physical environment

of climate, soil groups, and biomes with the human dimensions of politics, social structure, culture, and consciousness expressed by the idea of regionalism. This concept of region has also been referred to as a bioregion [7] or large landscape [8].

Regional, multi-stakeholder collaborations, which we will refer to as “regional environmental planning and design” (REPD) efforts, often arise from the identification of a particular resource problem, such as the cases reviewed in this paper concerning the management of wildfire, water quality, and biodiversity. However, such natural resource issues are inevitably intertwined with other social, economic, and environmental considerations. Given this complexity, large spatial scale, and organizational diversity, it is no surprise that REDP efforts face challenges in acquiring, managing, and sharing the information they need. A 2017 survey of 130 large landscape collaboratives found that “sharing and managing data at large scales and across jurisdictions” was among their major challenges, and that the top three tools and strategies provided by collaboratives were (1) facilitating strategic conservation planning, (2) information sharing, and (3) coordinating activities of partner groups [9]. In 2021, the United States president and leading federal agencies initiated efforts to support and align locally-led initiatives with the goal of conserving 30% of US lands and waters by 2030 [10]. This high-profile recognition of REPD initiatives increases the urgency for developing tools that can support and integrate REPD data and initiatives across scales and information silos.

Spatial data and analysis are critical to large landscape management, particularly in the recommended procedures related to assessing current landscape conditions and the collaborative spatial design of plausible future scenarios [11,12]. Until relatively recently, most spatial decision support activities were being carried out as desktop applications, since tools for sharing GIS functionality on the web were limited and challenging to implement. Now that the Internet and online mapping tools are widely available and easier to use, there are new opportunities for making such data and models more findable, accessible, interoperable, and reusable (FAIR) [13].

One of the most ambitious strategies for FAIR is Tim Berners-Lee’s (best known founder of the World-Wide Web) vision for a “semantic web” [14]. The underlying process relies on tagging information on the Internet with defined vocabularies for distinct entities, as well as for explicitly-identified relationships between entities. These standardized identifiers allow users to browse and search for specific classes of objects and their relationships, and they also allow computers to do basic reasoning, which improves search results and the ability to answer queries. Vocabularies of entities and relationships, akin to data schemas, are commonly referred to as “ontologies.” Recent advances in related knowledge management technologies, such as linked open data and knowledge graphs, have built on these ideas to provide new levels of information access, such as enhanced Internet search, e-commerce, and digital assistants (e.g., Apple’s Siri, Amazon’s Alexa, and Google’s Assistant) [15,16]. One example of the application of these technologies in the spatial analysis domain is the Spatial Decision Support Knowledge Portal [17,18].

In this paper, we use brief case studies to examine the information needs of three REPD efforts. Building on the theme of this special issue on “geospatial open systems,” we examine how semantic technologies, with knowledge graphs in particular, could extend the concept to “open knowledge networks” (OKN) that integrate both spatial and aspatial information across complex organizational networks. In Section 2, we describe the case studies and our approach to information needs analysis. We synthesize these needs across cases in Section 3, and in Section 4, we describe the design of an ontology schema to meet these needs. Section 5 describes how the REPD teams added data to the knowledge graph and provides examples of users’ queries. The final sections provide a discussion of our findings in relation to past and potential future work, and our summary conclusions.

2. Case Studies Needs Assessment

We identified three regional environmental planning and design case studies selected to provide diversity in their primary issue of interest. One group focused on water quality,

a second focused on biodiversity, and the third focused on wildfire. We identified the specific case studies based on where the authors had ongoing experience with and ability to characterize the needs of each REPD process. Such landscape collaboratives can be quite diverse [9], so we can make no claims as to the representativeness of this sample. However, by definition, all involve inter-organizational communication and the use of spatial environmental data.

For each case, we compiled information around the following themes, which were informed based on the authors' experiences and a review of associated documents:

1. What is the structure of the multi-stakeholder network and the problem they are addressing?
2. What are their knowledge discovery, sharing, and usage needs, particularly convergent ones across scales/sectors/roles/disciplines (we refer to these as "need-to-know questions and concerns" or NTKQC)?

2.1. Case Description: Water Quality in Puget Sound, WA

The Puget Sound region of Washington State is the second largest estuary by area and the largest by water volume in the U.S. It has recently received national attention for the plight of endangered orca whales, whose food supply of salmon is believed to be critically impacted by water quality. In turn, marine water quality is affected by large-scale climate change processes in the oceans and on land as well as the terrestrial water quality impacts of human activity.

To address this issue, Washington State has created the Puget Sound Partnership (PSP), which is composed of 750+ industry, government (local, state, national, international), not-for-profit, academic, tribal, and other stakeholder organizations [19]. The PSP has a wide variety of organizations involved and a fundamental need to synthesize information across ownerships and scales. They support the concept of integrated water resources management (IWRM), which is defined by the Global Water Partnership [20] as "a process which promotes the coordinated development and management of water, land and related resources, in order to maximize the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems". However, processes and tools to support IWRM are not yet widely available or adopted.

The PSP coordinating structure consists of the Leadership Council, the statutorily designated regional salmon recovery organization for Puget Sound, which together with the Tribal Management Council oversees decision making by the executive CEO [21]. Advising the executive and councils are the Ecosystem Coordination Board and the Science Panel, and in matters of Salmon Recovery, the Puget Sound Salmon Recovery Council. These boards guide the PSP and its many collaborator organizations.

PSP encourages convergence across disciplines and scales through a number of initiatives. They coordinate measurement and monitoring by developing a common set of indicators. They help coordinate research through a Science Panel, which consults with universities, agencies, and NGOs to develop 5-year work plans that identify key research objectives and criteria for prioritizing research projects. They also coordinate funding through the development of an Action Agenda [22] and through a cooperative agreement with the EPA. They work with Strategic Initiative Lead groups to manage subawards funded from the EPA's National Estuary Program fund. These subawards are chosen annually and are open to all organizations who propose projects working on Puget Sound recovery.

Based on the case study leads' work with the PSP and a review of available documentation [22–25], we identified two general types of information needs: (1) identifying indicators to assess and report on ecosystem conditions, progress toward goals, and the effectiveness of strategies and action, and (2) improving communication between various actors and constituencies, including scientists, decision makers, and the broader public.

The PSP has developed an extensive set of indicators, referred to as the "Vital Signs", covering their six broad goals [26]. However, they still face challenges in finding needed data, collating the diverse data which do exist, and using these indicators to evaluate

and prioritize actions. The PSP maintains an extensive online database related to their activities, including sections for indicators, activities, organizations, programs, and funds (<https://www.pugetsoundinfo.wa.gov/>, accessed on 2 July 2021). Some indicators include spatial data displays from an ArcGISOnline service, while others include static links to other data sources (e.g., <https://data.wa.gov/>, accessed on 2 July 2021). There does not appear to be any comprehensive catalog of these GIS layers. A related “Spatial Hub” website for the PSP only includes four datasets and nine mapping apps (<https://data-wa-ppsp.hub.arcgis.com/>, accessed on 2 July 2021). There has not been any formal data-sharing approach implemented across the Puget Sound, but a PSEMP Data Coordination Workgroup has been launched in 2019 [27]. Many of the indicators depend on biophysical modeling, so a related challenge has been tracking and combining the outputs of such models. The latest PSP Science Work Plan calls for coordinating the “... production and use of interdisciplinary research that explores and emphasizes the truly integrated nature of socio-ecological systems and multi-scalar dynamics, processes, and feedbacks between and within human and ecological components of the system” [23]. To connect these indicators with actions, the modeling processes underlying the Vital Signs indicators are being captured as “results chains,” which are conceptual models connecting actions to changes in indicators. However, there is not yet any formal pathway identified to operationalizing this process.

Connecting indicators to effective actions also requires communication between diverse actors and constituencies, including scientists, decision makers, and the broader public. The PSP Science Plan calls for invigorating interactions among scientists and decision makers, communicating findings clearly to diverse audiences, and improving the incorporation of indigenous knowledge. These communication challenges have been identified both within as well as between organizations.

2.2. Case Description: Wildfire Management in Santa Barbara County

The Cachuma Resource Conservation District (CRCD) is a quasi-governmental organization and part of a statewide network of resource conservation districts (RCDs). Their mission is to promote land ethics that result in the long-term use of natural resources while protecting and enhancing natural habitats. They were funded by the California Department of Conservation, via the Coastal Conservancy, to develop a Regional Priority Plan to reduce wildfire risk and improve forest health for Santa Barbara County. The Plan is meant to help with the mapping, prioritization, and planning of risk reduction projects. Reducing wildfire risk is a convergent problem because the system of people and organizations involved is complex and diverse: from households and neighborhoods to non-governmental organizations and volunteer groups to businesses and investors to agencies and elected officials at all levels of government. The CRCD leadership consists of a Board of Directors, an Executive Director, and contractors. They primarily coordinate between stakeholders through group meetings and hiring contractors to co-design products. Meetings include a diverse set of experts and stakeholders, and products, such as synthesis reports and educational videos, demonstrate how these diverse views are channeled into convergent problem solving.

The CRCD and associated consultants held a series of meetings in late 2019 and early 2020 to identify information needs. Through these meetings, combined with documentation from the CRCD and partner websites [28–30], we co-identified three general types of information needs: (1) identifying and prioritizing areas for wildfire risk reduction efforts; (2) identifying and prioritizing actionable projects that mitigate wildfire risk, build community capacity, and increase wildfire and climate resilience; and (3) identifying the primary network of people and organizations that should be involved for any particular area and/or project.

To identify priority areas, the CRCD has partnered with the Conservation Biology Institute and is using their online mapping portal Data Basin [31] as a platform for gathering and sharing data. The project is also using the environmental evaluation modeling system

(EEMS), which is a spatial multi-criteria model that plugs into Data Basin with an associated graphical user interface [32], to prioritize parcels and properties for fire risk reduction. There is a need to network with local experts, stakeholders, data managers, and decision makers to design the hierarchical, multi-criteria model such that it incorporates all the criteria of interest in the region for which adequate data are available, and to combine and weigh the data appropriately.

To identify and prioritize projects, the team interviewed more than 40 stakeholders and decision makers to develop a list of projects and had follow up conversations to rank priority. However, the actors in this system often do not know of each other's activities, leading to the potential for duplicate or conflicting actions. Making the relevant system of actors and programs more readily visible and understandable should yield significant gains in individual, agency, and collective system knowledge and efficacy in this critical work. CRCD is looking for tools that measure levels of connectivity, communication, coordination, and collaboration between and amongst the "collaborative web" of actors in the system, and track changes over time. They would like the tools to aid collective efforts to improve the system, help in decisions about where to direct resource investments, and how to implement those investments to reduce risk.

The team has a rich spreadsheet of important people and organizations involved in various aspects of fire risk reduction, a rich spreadsheet of projects, and the spatial analysis of priority locations for risk reduction in general, for risk reduction in open space, and for prescribed herbivory. The team lacks the capability to integrate the people/organizations database with the projects database and the spatial prioritization information and has identified this as a major need.

2.3. Case Description: Biodiversity Conservation in the Cascades to Coast Region

The Cascades to Coast Landscape Collaborative (CCLC) of western Washington and Oregon is a self-directed collaborative that brings together a multi-jurisdictional group of natural resource partners, including agencies, NGOs, large and small private landowners, tribes, farmers, land trusts, and smaller municipalities. The two primary intentions of the CCLC are to collectively create a landscape-scale design that best supports shared ecological, economic, and cultural values as they relate to biodiversity conservation, sustainable working lands, and ecosystem services, and to provide a forum for bringing people together around shared goals, shared information, and shared action (<https://www.ctoclc.org>, accessed on 23 May 2021). The CCLC originated as a project of the North Pacific Landscape Conservation Collaborative.

CCLC has adopted a landscape-scale conservation planning approach based on Campellone et al. [12]. This approach has four key attributes: convening stakeholders, assessing landscape condition, spatial design, and strategy design. The purpose of strategy design is to translate science products into mutually reinforcing strategies that identify stakeholder roles in fulfilling a shared vision for the landscape. These four attributes underpin the goals of the CCLC, which are (1) to create and house spatially enabled, science-based tools that can advise and support local management plans and conservation priorities to facilitate a resilient landscape, and (2) to collaboratively develop strategies to design and manage resilient landscapes for people, ecosystems, biodiversity, working lands, and communities.

The CCLC coordinating structure consists of a facilitator/coordinator, GIS, web design and engagement contractors, and a volunteer and in-kind leadership team of knowledge experts from multiple agencies, organizations, and owners of working lands such as forests. The CCLC encourages convergence across disciplines and scales through frequent partner engagement by holding workshops in different locations across the CCLC's large bio-region, webinars, and a resource-rich website for learning, discussion, and shared output. They initiated an ArcGIS Online tool to make their regional spatial modeling results accessible publicly, and they launched a simple, web-based search tool to enhance engagement by

small forest landowners with state and federal financial and technical incentive programs that support conservation.

We identified three categories of information that were important for the Cascades to Coast Landscape Collaborative: (1) information about important areas to conserve or restore [33], (2) information on strategies, opportunities, and tools for conservation implementation [12], and (3) information about who is available to engage with planning or implementation [34,35].

A major challenge for identifying priority areas to conserve is that detailed information that supports productive dialogue is spread across what traditionally have been information silos [36]. For example, state wildlife agencies house spatial data on important fish and wildlife habitats, while county governments house data on locations of rural resource lands, while land trusts have a sense about the desires of the local land-owning population. CCLC priority needs include combining disparate data layers while also providing a launch pad for public engagement about why these sites are important. For example, when a site is identified as important for “habitat connectivity”, information needed to make the case for protection of this value (connectivity) is based on specific information such as which species, at which life history stages, require this connectivity function. There is a danger when rolling up too much information into a simplified model of spatial priorities. People want to know “why” and be able to query the supporting information and see specific criteria mapped. Showing “why” along with the “what/where” can create a catalyst for interested parties to join forces and take collective action around particular geographies. Do we need mature forests in that area? Can connectivity be maintained if timber is extracted? If the land is converted to agricultural or low-density residential, or if roads or energy-related infrastructure are added, would the site still function for connectivity? When information is readily available on why a site is important, people can assess the threats to those values and devise strategies to address those threats.

A priority of the CCLC is to look for and support common interests among different initiatives to enhance the potential for new partnerships. Much of the land in the CCLC bioregion is privately owned; therefore, landowner engagement by public agencies and NGOs is very important. The CCLC built a web-based clearinghouse tool for landowners to more seamlessly find government incentive programs. However, this tool, although very useful to landowners, is itself currently siloed and not readily linked to other information. Regardless of the user, whether a forest landowner or agency scientist, queries that support landscape design must be simultaneously customizable, granular, and integrated, providing seamless access across information silos.

3. Case Studies Needs Synthesis

The objective of the research reported herein was to focus the development of our open knowledge networking tools by synthesizing needs across the three case studies to identify the most common, shared decision support needs. Table 1 summarizes the high-level information needs described by the case studies in the previous section. Information about people and organizations was a priority for all the cases, which was followed by an interest in identifying priority areas for work.

For a more detailed analysis of case study needs, each of the case teams generated a list of 20–30 need-to-know-concerns and questions (NTKQC) with the idea that these could be eventually transformed into “competency” questions for the knowledge graph (Table S1: Case Study Questions and Synthesis Clusters). Competency questions are a common way of specifying knowledge graph (KG) needs and testing whether the KG can meet user needs [15].

However, for prototype KG development, we did not want to attempt to answer all questions but rather to begin by focusing on the top common needs shared among the cases. We identified these shared needs through a two-step process. First, case leads were asked to synthesize their question lists into a shorter list of about five more general needs for each case (Table 2). Then, the project and case leads discussed similarities and came up

with a more abstracted set of categories to define these synthesized needs (Table 2 Synthesis Categories). A primary category was assigned to each need, and a secondary category was assigned where a need fit within more than one. Results based on these four categories are described below.

Table 1. High-level information needs from the case studies.

Case Study	Information Needs			
	People and Organizations	Indicators	Priority Areas for Work	Strategies, Methodologies, and Tools
Water quality in the Puget Sound	✓	✓		
Wildfire management in Santa Barbara	✓		✓	✓
Biodiversity in Cascades to Coast	✓		✓	✓

Table 2. Synthesized user needs from detailed case study questions lists.

Case	Synthesized Need	Synthesis Category 1	Synthesis Category 2
Water	Need to develop capacity and coordinate efforts to assess and report on ecosystem conditions (and progress toward goals) and the effectiveness of strategies and actions	Information on where/what	Information on why/how
Water	Need to collaboratively develop an interdisciplinary science enterprise to support Puget Sound ecosystem recovery	Communication within/between roles	
Water	Need to develop and analyze alternative future scenarios to explore and express desired futures and evaluate trade-offs among possible approaches	Information on where/what	Information on why/how
Water	Need to improve and invigorate interactions between scientists, managers, and decision makers	Communication within/between roles	
Water	Need to engage the public, especially on tough trade-offs	Communication within/between roles	
Water	Need to improve incorporation of indigenous knowledge into science and monitoring efforts	Communication within/between roles	
Biodiversity	Need to combine disparate data layers to identify where are the important areas for conservation, and where are the important areas for working lands	Information on where/what	
Biodiversity	Need tools to help diverse audiences understand how and why sites are prioritized and engage in the decision and implementation processes	Communication within/between roles	Information on why/how
Biodiversity	Need methods to identify and prioritize what strategies, opportunities, and tools are available to a diversity of users for implementing this landscape design	Information on why/how	
Biodiversity	Need to identify who is available, interested, and able to support the implementation of actions	Information on who	
Wildfire	Need to make the system of actors more readily visible and understandable	Information on who	
Wildfire	Need to measure levels of connectivity, communication, coordination, and collaboration between and amongst key actors in the system and track changes over time	Communication within/between roles	
Wildfire	Need tools to help guide risk reduction efforts and direct resource investments to particular places	Information on where/what	
Wildfire	Need methods to connect the people data with the environmental data (e.g., the network of people and organizations that should be involved for actions on any particular parcel)	Information on who	Information on where/what

3.1. Information on Where/What

All of the collaboratives had information needs relating the where and what of resources and activities. The biodiversity and wildfire cases share a common need for identifying priority areas for action, such as “where are the important areas to focus on for conservation” and “identify particular parcels in a countywide extent that are priorities for wildfire risk reduction efforts.” The water case needs were more abstract, but a number still implied the need for where/what information (indicators, information for planning, future scenarios).

3.2. Information on Why/How

Information on the why and how of conservation actions was related to two need statements in each of the water and biodiversity cases. Both of the needs expressed by the water and one of the needs expressed by the biodiversity case were more instrumental (identifying strategies and trade-offs). The biodiversity case brought up a more communicative need, which was for transparency decision making in order to facilitate the buy-in of stakeholders.

3.3. Information on Who

The need to make the system of actors more readily visible and understandable was most strongly expressed in the wildfire case, along with a need to connect these people data with the other resource data they are using for prioritizing wildfire treatments. The biodiversity case also had a need to identify relevant actors but expressed more specifically related to the implementation of conservation actions. The water case needs did not explicitly include this tag, but with four of their six needs involving communication, this need for information on actors could be inferred.

3.4. Convergent Communication

This category was the most frequent, being applied to six of the 14 expressed needs. As mentioned above, it was particularly important in the water case. This importance may be due to the fact that the water case has the longest history, most actors engaged, and there were many links to state and federal policy processes. Both the water and biodiversity cases explicitly mention the need to communicate between the roles of scientists, decision makers, and stakeholders. The wildfire case expressed an additional need to track these types of communication over time.

4. Knowledge Graph Schema Development

As described in the Introduction, our working hypothesis was that semantic technologies, and knowledge graphs in particular, could help to address some of the challenges faced by REPDs, including integrating information across complex organizational networks and across an array of tools developed for narrow (often disciplinary) applications.

There is still some debate around the definition of knowledge graphs, but a recent review of the field defined them as “a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities” [15]. A variety of graph data approaches, data structures, and query languages have been developed. For this project, we chose to follow the World Wide Web Consortium’s Resource Description Framework (RDF) and the Ontology Web Language (OWL) standards [37,38]. Here, data are fundamentally stored as “triples” consisting of a subject–predicate–object relationship, for example: ex:Jane Doe ex:is member of ex:WA Dept. Natural Resources. Each of these elements is also uniquely identified by assigning a prefix (ex: in this example) that references it to a specific data store, so ex:Jane Doe is distinct from ey:Jane Doe. These prefixes are referred to as Uniform Resource Identifiers (or URIs), and for linked open data on the web, they are generally resolvable URLs. So, for example, the core RDF schema elements are referenced to <http://www.w3.org/1999/02/22-rdf-syntax-ns#> (accessed on 4 January 2021), with an

abbreviation to ‘rdf:’ usually set in the header of the dataset. Similar to object-oriented programming, entities are often assigned to broader classes, which then have certain types of properties (ex:Jane Doe rdf:type foaf:Person), where foaf:Person has properties such as ‘has first name’, ‘has phone number’, etc. Perhaps the most unique aspect of this approach is that data and metadata are effectively combined, enabling more precise searching and reasoning over such data stores. The SPARQL query language standard (similar in purpose to SQL for relational databases) has been developed to facilitate querying across such data stores [39].

In this section, we describe three initial steps in KG schema (ontology) development: (1) identify the priority subset of user needs to address; (2) conceptually define a set of entities (classes) and relationships needed; and, (3) operationally define the ontology framework (incorporating existing ontology elements where appropriate).

All the case studies identified the need for information about people and organizations, and two of the four synthesis needs concerned people and communication (see Table 1). Thus, we chose these types of information as the initial focus of our work. In future versions, we hope to address the additional needs related more specifically to where/what (i.e., priority areas for work) and why/how (i.e., implementation strategies). Starting with this priority, the case study leads worked with knowledge engineers to identify KG schema entities and relationships that would support these needs. The initial list of classes identified included Person, Organization, Program, Project, Report, Tool, Dataset, and Indicator. Additionally, it was seen as important to be able to search these by Region and Knowledge Domain (Table 3). At the heart of these classes are the common needs of all our collaborators to know who is doing what work, what are the products of that work, on where the work is focused, and how resources are organized and funded to do that work. These types of information do not by themselves address the need for “convergent communication” between roles, but we believe that providing more accessible information on the activities of the various actors is a first step in that direction.

Table 3. List of conceptual entities needed for the knowledge graph.

Class	Description
Person	An individual person with important links to organizational entities, products, and domain expertise
Organization	A group of people with a commitment to a shared structure and purpose
Program	An organizational structure for the resourcing and organizing of ongoing work
Project	A temporary organizational structure for accomplishing a particular objective
Report	A document describing the results of some investigation or effort
Tool	A codified technique for accomplishing a particular objective
Dataset	A named collection of data, usually containing only one type of data
Indicator	A piece of information designed to be a measure of the state of some phenomenon
Region	A particular spatial area
Knowledge Domain	A category for describing a body of related information

Given this list, we researched elements from other existing ontologies to the extent practicable both to jump start our own schema design and to improve compatibility with other linked data. Entities representing people are found in many existing ontologies, including one of the best-known examples, “Friend of a Friend” (FOAF) [40]. However, neither FOAF nor its integration into broader ontology frameworks, such as VIVO [41], W3C Organizations [42], and schema.org [43] express the richness of relationships to organizations, projects, and what their collaborative work produces—for example, datasets and reports—needed by our users. The best existing match we found was the Persons–Projects–Organizations–Datasets (PPOD) ontology developed at the University of California–Davis to characterize a network of entities for describing food production in a locality (a “foodshed”) [44–46].

4.1. PPOD

The PPOD ontology contains 112 classes, 75 object properties, and 8 data properties [45]. It draws upon a number of existing ontologies including FOAF, VIVO, AgriVIVO [47], the Relation Ontology [48], an ontology describing sustainability issues and associated indicators [49], and an ontology describing expertise [44]. As reflected in the title, central classes include Person (the who), Project (a way resources are organized), Organization (a way people are organized for work), Program (a way that projects are organized), and Dataset (one example of what work produces). Additional classes related to Person include Source (primary source for the information about an entity) and Role (a class allowing for the enumeration of the details of the relationship one entity has to another). Taxonomies were developed for describing organization type and activities, project type, and program type. Most of the 112 classes in PPOD are subclasses of these major classes. For example, the Organization class is subdivided into 41 subclasses, examples being Institute, Museum, and Library. Most of these organization subclasses were imported from the VIVO ontology. PPOD only contains only three additional subclasses that are locally defined (Consultancy, Media Organization, and Commodity). Currently, PPOD contains a general geospatial class named Geographical Region, from which two subclasses are defined to hold the information for California counties and the U.S. Department of Agriculture ecoregions. Organizations and other classes are linked to the spatial entities in which they operate (rather than within which they are located). These spatial units are simply represented as named entities; however, no spatial coordinates are stored or used.

4.2. Extending PPOD to PPOp (People, Projects, Organizations and What They Produce)

PPOD provided a strong initial schema for PPOp, but from our review, we identified two additional needed classes: Tool and Report. Additionally, four existing classes needed significant extensions: Program, Dataset, Indicator, and Geographical Region. These modifications are described in the following sections, and the most important entity types and relationships for PPOp are shown in Figure 1. The prototype ontology is available in a public repository on Github (<https://github.com/SDS-OKN/PPOp/>, accessed on 22 November 2021).

4.2.1. Program Class

Significant work in the world is ongoing, and the resourcing and organizing of ongoing work is very often delivered through programs. A program may support multiple projects, create tools, and generate multiple datasets over time. Often, a program uses certain policy tools to achieve its objectives (e.g., authority, incentive, capacity, persuasion, learning [50]). Organizations may fund and/or participate in a program. A program may even support other programs. Program classes are found in the existing ontologies for schema.org and VIVO, which we have incorporated and built on to meet our use cases. A full list of properties and sources can be found in the Supplementary Materials (Table S2: PPOp Program class properties). Relating to the synthesized need for “what/where” information, the Program class has a property `has_subject:Knowledge_Domain`, while “where” information is captured by the Region class (see further description below).

4.2.2. Tool Class

Environmental planning and design activities often involve analytic work, which gets implemented through the development of information tools. For this work, we focused on data transformation tools: tools that take input (parameters, models, datasets) and generate new datasets, as well as data visualization and exploration tools. The SDS Ontology Framework [17], which we will federate to in a later phase of this project, contains both a wider definition of tools (decision process tools, meeting tools) and a deeper decomposition of tools into methods and workflows. The Tool class described here is a “stub” for that more complicated Tool class in the SDS ontology framework. Typically, the tools we focus on enable us to estimate the values of Indicators or simulate system processes, e.g.,

ecohydrology or wildfire spread. The Tool class links to other classes through properties such as `has_subject:Knowledge_Domain` and `has_output:DataSet`.

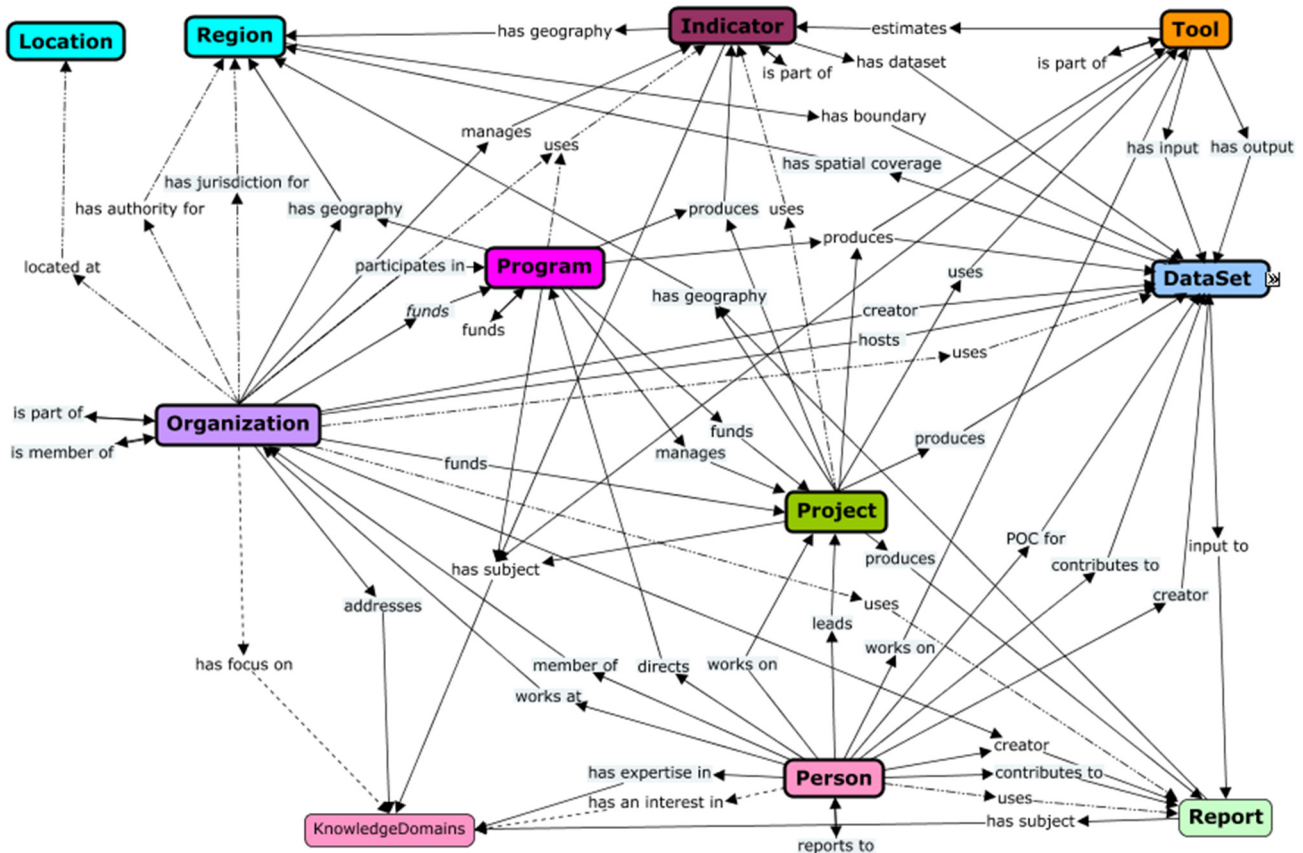


Figure 1. Key object properties of the main classes in PPOp. Solid lines relate to scientists creating information, dashed lines relate to decision makers and managers using information, and dotted lines relate to stakeholders having an interest in issues.

4.2.3. Report Class

Many of the results of work activities relevant to REPD's are written up in reports. Reports may be either private (restricted) to an organization or publicly distributed. However, even if public, reports may be difficult to find, since they are typically not captured by the major commercial or academic literature indexing services, such as Web of Science or Scopus. Such reports are usually only available on the website of the organization that produced them. For our Report class, we started with the base Report subclass of the Document class from the well-known BIBO ontology [51]. This subclass provides basic attributes, such as title, creator (author), date, etc. To this base, we added a number of links to our other core PPOp classes. Reports can be generated by Programs and Projects. Many Persons contribute to a Report, and a Report can be about results derived using specific Tools and captured in specific Datasets. A Report may also be linked to one or more Regions and Knowledge Domains.

4.2.4. Dataset Class

Data about biophysical and socio-economic aspects of landscapes help REPD's understand and make decisions about them. We built on the Dataset class in the PPOD framework (in turn sourced from the VIVO ontology), which already included potential connections to Persons and Organizations as creators, a geographic focus, and subject area. We added links to the Program and Project classes as creators and Reports and Tools for which Datasets can be inputs or outputs. We added more detailed properties for spatial and temporal resolution. Datasets are also increasingly available via web services, so we added a data property `has_webservice`. For instance, a specific USGS hydrologic unit can be represented

by a static feature class, and we can provide URLs to where it can be downloaded, but it can also be provided via a dynamic web service endpoint hosted on ArcGIS Online. The following end point returns the boundary of the HUC 8 watershed “Hood Canal” in the Puget Sound: <https://hydro.nationalmap.gov/arctgis/rest/services/wbd/MapServer/4/query?text=Hood+Canal&geometryType=esriGeometryEnvelope&outFields=Name&returnExtentOnly=false&featureEncoding=esriDefault&f=json> (accessed on 6 January 2020). Such web services can always provide the latest version of the data and avoid the need to store these data locally.

4.2.5. Indicator Class

Particular metrics, often referred to as “indicators,” are often used to understand the condition of the world in order to plan where work needs to be done and how to prioritize resources for work. Decision support projects identify significant indicators, what direction of change in an indicator represents improving conditions, and how to synthesize movements in the values of multiple indicators to determine aggregate progress toward desired conditions. Indicator values can be monitored and recorded in Datasets and future values simulated by means of Tools. By connecting Projects to their ability to move Indicators in a positive direction, we can address why/how needs. We can explain why those projects should be funded, how they will affect each indicator value and the change in values anticipated which, if the project is funded and implemented, can in the future be validated against monitored data.

4.2.6. Knowledge Domain Class

Many of the need-to-know questions and concerns involved the filtering of the classes by some form of knowledge domain or issue. The Knowledge_Domain class serves a similar function as the PPOD Issue class. Many of the core classes (Person, Project, Datasets, Indicators, Reports) are associated with Knowledge_Domains through the has_subject relationship. These Knowledge_Domain lists grew organically as we added instances of classes in the Santa Barbara and Puget Sound use cases. Knowledge domains cover a wide variety of themes, many of which could be described through their own ontologies. In this first iteration, these domains are handled by a single list of terms, but future development could link to domain-specific ontologies and knowledge graphs. Some examples of such ontologies that may be of particular relevance to stakeholders in our current use cases include water [52], biodiversity [53], and wildfire [54].

4.2.7. Region Class

It was clear from the questions we gathered from the case studies that many questions involved reasoning about geospatial qualities: what are study areas for reports and datasets, how to find individuals with particular geographic expertise or interests, and where do projects, programs, and organizations operate. Regions can have their definition provided by a wide set of considerations—physical, jurisdictional, political, etc. As the number of types of regions increases, it becomes challenging to tag each of the related class instances with all the potentially associated regions. Given that regions are areas in the world that can be described or at least approximated by boundaries made up of linear polygons, GIS handles these relations through spatial overlays. A few spatial extensions to RDF and the associated SPARQL querying language have been developed. The most common geospatial information in linked data is simple latitude/longitude point locations defined through the informal Geo standard [55]. More sophisticated semantic approaches have been developed, including stSPARQL [56] and GeoSPARQL [57]. Although GeoSPARQL has only been implemented in some software platforms (see [58]), we aligned our ontology with it because it is an industry standard that seems to be growing in use.

In GeoSPARQL geographic objects are assigned to the class Feature, which has a list of standard properties, of which we have implemented hasBoundingBox (for generalized queries) and hasGeometry (for holding more specific geometries). Features can have the geometries of points, lines, or polygons (GeoSPARQL does not currently handle raster data).

The PPOD ontology included a Region class from the Geographical Entity Ontology (<http://www.ontobee.org/ontology/GEO> accessed on 22 November 2021), which is part of the larger Open Biological and Biomedical Ontology (OBO) Foundry (<http://obofoundry.org/> accessed on 25 November 2021). In PPOp we simplified the name to Region and linked it to the classes Person, Organization, Program, Project, Report, Indicator, and Dataset through the predicate `has_geography`. In turn, a Region can be declared as a subclass of Feature and then link to a specific geometry object via the GeoSPARQL predicate `hasGeometry`. Organization is also related by predicates `has_authority_for`, `owns`, and `is_active_in`. Listing 1 presents an example of how a particular project is represented as a Region in our RDF data. However, these spatial features are not required for name-based queries, so a user adding a Region instance need only focus on the identification/definition and not get side-tracked by GIS data requirements. Once a region is fully attributed with a boundary dataset, it will support spatial as well as logical inference when resolving queries.

Listing 1. Example RDF Turtle code for describing a Project’s spatial geometry.

```
@prefix: <https://github.com/SDS-OKN/PPOp/blob/main/ppop.owl#>.
@prefix geo: <http://www.opengis.net/ont/geosparql#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

# Declare our Region to be a subclass of the GeoSPARQL Feature class
:Region rdf:type rdfs:Class;
rdfs:subClassOf geo:Feature.

# Create an entity for a project with a link to a geometry object
:ProjectX rdf:type: Project;
rdfs:label "Project X";
geo:hasGeometry: geom_ProjectX.

# Create the geometry object for the feature as a series of xy coordinates
: geom_ProjectX rdf:type: Region;
geo:asWKT "POLYGON (48.496 -124.673, ...)""^geo:wktLiteral.
```

5. Knowledge Graph Implementation

Operationalizing the knowledge schema developed in Section 4 involved preparing data, checking its functionality, managing data and schema changes over time, and demonstrating their utility through use cases.

5.1. Data Population

Populating the KG schema with actual data records (often referred to as “instances” in semantic approaches) on people, projects, and the other schema entities entailed some challenges. The basic needs identified for a data entry system for the REPD efforts were (1) a user interface simple enough for users without semantic data training, (2) easily accessible across multiple organizations, (3) low cost, and (4) compatible with the RDF/OWL open data standards. Secondary needs included secure user logins and versioning. While a considerable number of graph database platforms and vendors exist, we did not find any off-the-shelf software solution that met these basic needs. Two of the closest solutions included WebProtege [59,60], a free web-based version of the Protege editor, and Gruff [61], a free web-based interface to the Allegrograph triple store. Both met all the needs except that their interfaces are too complex for nontechnical users (late in the process, we discovered that WebProtege does include an option for customizable forms for data entry, but the configuration process was deemed to be too complex to implement at our late stage of prototyping). A third possible option is the Semantic Mediawiki (SMW) [62] extension to the popular Mediawiki software (which powers Wikipedia). SMW met all the requirements, notably with a relatively simple user interface, but it does not support the

RDF/OWL and SPARQL standards without the setup of an additional RDF-compatible database. In addition, as with the Gruff option, SMW requires the setup and maintenance of the software on a hosting server, which did not seem feasible for our REPD collaborators.

The solution implemented to support the case study groups entering instances to the knowledge graph was to set up a series of Google spreadsheets that followed the ontology structures in a tabular format. Google's document system enables easy web-based access and user permissions to accommodate users across different organizations, and the spreadsheet interface is simple and familiar to most users. In the spreadsheet, each class was represented by a tab and each class property is represented as a column. In order to facilitate class–predicate–class (subject–predicate–object) linkages, we set up class range columns using lookup functions to the sheet of the object class, which were implemented through Google sheets data validation services. For class–predicate–data relationships, where the data was to be from a controlled vocabulary, we used the same approach but validation was against lookup ranges containing the controlled vocabulary. Where one too many relations are needed, we used commas to separate range values and updated a common Google sheet tool to facilitate the selection of multiple items from lookup ranges. Where a predicate would have different instances differentiated by time, for instance, a Person works at an Organization and it is important to know when they worked where, we implemented "roles" by introducing sheets such as PersonOrg where start and end dates and titles could be added to a person's tenure at an organization. Finally, identifiers and rdf:labels were automatically generated that were admissible in OWL representations of the KG.

Participants from the water quality and wildfire cases entered data for the classes described (the biodiversity case did not due to other priorities). For some of our geospatial entities, we were able to link existing web data services to the spreadsheet using macros. For example, the US Geological Survey provides their hydrologic unit code data (HUC, a hierarchy of watershed boundaries) via a web service [63], which is linked into our spreadsheet, so users can select an HUC of interest from a drop-down list to populate the geographic data for hasBoundingBox and hasGeometry. We created a Python script to take the entries in the Google spreadsheet and convert them into an RDF file in Turtle format [64]. This script has a series of dictionary entries describing the predicates represented by each column in the spreadsheet: for instance, whether the predicate represents a data property or an object property, the URI for the predicate, and the text label to be used for the predicate. The script iterates over each row in every sheet and populates an RDF file using the Python library RDFLib [65].

5.2. Schema Management

For local checking, editing, and querying of these RDF files, we used the free, open-source ontology editor Protégé [66,67]. For initial testing as a web resource, we used the Allegrograph engine [68] based on familiarity from previous work. However, Allegrograph does not support the GeoSPARQL standard for geospatial data, so additional testing was done using the GraphDB engine.

As with any information systems project, user needs will evolve over time, and procedures are needed to modify the system. The wildfire and water case studies that moved to gather instances related to work in those areas started with a common Google sheet template of classes (described above) but added new classes (sheets), new properties (columns to existing sheets), and extended controlled vocabularies as necessary. We established a bi-weekly steering committee to track and guide these extensions, and we ultimately decide on a synthesized set of classes and predicates we would import into a common prototype KG. We found that synthesis was best achieved by replacing very specific predicates ("has topical focus") with more generic predicates ("subject"), which makes resolving conflicts across cases easier, at the cost of specificity. As the collaboratives move from the initial prototype to fuller implementation, we would expect each to manage their own datastore, with the flexibility to modify the schema based on their particular

needs. Assuming these datastores are web-accessible, it would still be possible to query across them, as we describe in the next section.

5.3. KG Query Use Cases

In order to demonstrate the utility of the PPOp knowledge graph, we present a number of use cases below with specific competency questions. Queries on RDF data are supported by the SPARQL query language specification [39] and the GeoSPARQL extension [57] for geospatial queries. For the first two use cases, we include SPARQL examples. SPARQL has some similarities to SQL for relational databases but tends to focus more on matching triple patterns (subject–predicate–object). Items beginning with “?” denote a variable that returns whatever entities match the rest of the triple pattern (in the simpler examples, we use ?s–?p–?o for subject–predicate–object). Then, these variables may be used in other triple matches and as elements to return in the result set (those items following the SELECT statement). As discussed in Section 4, each entity in the graph is represented by a URI, and for ease of specification, the part referring to the general ontology is usually assigned to an abbreviation using the PREFIX statement. Here, we have assigned our PPOp URI to the prefix “ppop”. For brevity, we provide example output only for the first query.

One of the most basic queries a user might be interested in is to find all the data in the graph about a particular entity: for example, the Person “Jane Doe”. In a relational database environment, this query might involve a number of table joins (requiring the knowledge of the tables and their joining parameters), but in the RDF data model, all the information about an entity was stored locally (roughly analogous to a single record), so this query can simply be performed as follows (with output for the query is presented in Table 4):

Table 4. Example output from the SPARQL query for all items directly related to “Jane Doe”.

?p	?o
type	NamedIndividual
lastName	“Doe”
homepage	“ https://www.linkedin.com/in/jane-doe-0848b121/ ”
type	Person
worksOn	Watershed Characterization project
holdsJobTitle	Scientist
worksAt	Washington Dept. of Ecology
has_expertise_in	Water_Quality

```
PREFIX ppop: <https://github.com/SDS-OKN/PPOp/raw/main/ppop.ttl#>
```

```
SELECT ?p ?o WHERE { ppop:Jane_Doe ?p ?o }.
```

Finding and tracking clusters of interest is important for collaboratives. For example, if a decision maker is interested in organizing a meeting around a specific topic, they would want to know about all the related PPOp entities. The flexible semantic data model also makes searches across different entity types easy. For example, one can find all the entities with an interest or expertise in water quality using the following query, which produces a single table incorporating persons, projects, programs, and organizations. The “|” character functions as the logical OR operator, and we use ORDER BY to sort the results:

```
SELECT ?type ?entity WHERE {
  ?entity ppop:has_an_interest_in | ppop:has_expertise_in ppop:Water_Quality.
  ?entity rdf:type ?type. } ORDER BY ?type ?entity.
```

If the organizer wanted all the persons (and just persons) associated with the projects-/programs/organizations returned in the previous query, the graph could easily trace these connections as follows:

```
SELECT ?person ?entity WHERE {
```

```
?entity ppop:has_an_interest_in | ppop:has_expertise_in ppop:Water_Quality.
?person ppop:worksAt | ppop:worksOn ?entity. } ORDER BY ?entity ?person }.
```

Of course, one of the main advantages of linked open data is that queries may be made across different information stores. After the prototyping stage, we expect that the collaboratives would maintain and grow their own data repositories, but it still would be easy to search across them by simply including their service locations in a query as follows:

```
SELECT ?person ?org WHERE {
  SERVICE <https://www.psp.wa.gov/sparql> {
    ?person psp:has_expertise_in psp:Water_Quality.
    ?person psp:worksAt ?org. }
  UNION
  SERVICE <https://www.sbcwildfireresilience.org/sparql> {
    ?person sbc:has_expertise_in sbc:Water_Quality.
    ?person sbc:worksAt ?org} }.
```

Using the Geosparql extension, spatial queries may also be made. A sample user question might be, “For a given project footprint, what datasets are available?” The query can be formed by first requesting the geometry associated with the project of interest (here denoted as ProjectX), then searching for datasets having geometries overlapping with the project geography (using the Geosparql spatial operator `sfOverlaps`). Note that the “;” character extends the same subject to further predicate-object queries, and the `geo:` prefix refers to elements from the Geosparql standard.

```
SELECT ?title, ?url
WHERE {
  :ProjectX ppop:has_title ?title; ppop:has_url ?url; geo:hasGeometry ?project_geo.
  ?dataset rdf:type: Dataset.
  ?dataset geo:hasGeometry ?data_geo.
  ?project_geo geo:sfOverlaps ?data_geo. }
```

6. Discussion

Our work built on the results of an earlier broad-scale survey of landscape collaboratives that identified “sharing and managing data at large scales and across jurisdictions” as one of their top needs [9]. To our knowledge, this paper provides the first more in-depth assessment of these needs in the peer-reviewed literature, albeit for a small but diverse sample. Other environmental management-related needs assessments in the literature have been centered on particular topics and/or types of users rather than the networked set of actors and interests in our place-based collaboratives. However, some common themes emerge, which our knowledge graph approach could help support. For example, in studies of biodiversity information users, two of the six broad priorities identified by Smythe et al. [69] included “information management strategies” and “communication and outreach”. More specifically, 90% of Steiner Davis et al. [70] respondents stated the importance of biodiversity search tools to their work, but only one-third of respondents said it was easy to find the information search tools they needed. In a survey of wildfire and fuels management professionals, Ryan and Cerveny [71] found related barriers of “lack of time” and “information overload.” KGs enable more specific searching, which should help filter out irrelevant information and reduce the time to finding pertinent information.

However, improving search is not a panacea for meeting users’ information needs. In the field of water management, Gober et al. [72] found institutional factors impeded collaboration and information sharing between land and water managers. Likewise, Rayner et al. [73] found political and regulatory impediments to new types of information use among water managers. Potential solutions to such challenges include the creation of boundary objects or organizations to help bridge institutional silos [74] and increasing the collaboration between scientists and managers for the co-creation of useful information [70]. While KGs cannot solve institutional barriers on their own, they have demonstrated

potential to integrate different information silos and could provide part of a platform to help producers and users collaborate on the production of usable information [16,75].

Another key finding we made in the process was that while the classes we added to extend the original PPOD helped address who/where/how questions related to work being done, we realized that we also needed to provide support for planning work—why a proposed project would be beneficial, to whom, who would fund it, and how can we know if it is effective? The Indicator class is a first step in this direction, and we look to incorporate the Funding, Research Administration, and Projects Ontology (FRAPO) [76] in the next iteration.

6.1. Available Technologies

Despite the conceptual potential for KGs to improve information development and use, we found a number of technical impediments to their deployment. A variety of ontology editors, many of them open source, have been available for some years [77,78], as have a few options for graph databases. However, available software for managing KGs appears to be too complex for most users who want to create or even just browse KGs. Compounding this complexity is the fact that separate tools are often needed for importing data, ontology editing, storage, visualization, etc. Hitzler [16] also noted this need for “easy-to-use and well-integrated tools supporting the whole process” in the semantic web field. However, software in this sector is evolving rapidly, and simpler, cloud-based options with low fees may provide a solution (e.g., [79,80]).

6.2. Semantic Data Expertise

Finding the necessary semantic data expertise is a second challenge. While REPD personnel seem to readily grasp the subject–predicate–object conceptual structure for graph data, actually generating these data, even using some form of dedicated editor, is considerably more difficult. Our REPD case studies all have personnel involved who have expertise in databases and GIS but none with semantic data experience, nor is this type of training generally available in college curricula. Two exceptions worth mentioning are open courses on knowledge graphs at Stanford [81] and the Hasso Plattner Institute [82]. With durations of 6–10 weeks, such courses likely require more of a commitment than is feasible for most REPD personnel. Some shorter self-directed training materials exist online, but they are not coordinated and vary considerably in quality and prerequisite knowledge required. Development of an open textbook [83] could provide a useful reference, but this would likely require dedicated funding for the development and maintenance in this rapidly evolving field (perhaps supplemented by some type of wiki crowdsourcing). The development of simpler, more integrated software with associated training materials may be the most realistic short-term solution to this bottleneck.

6.3. Data Population and Schema Management

Complex software and limited expertise create challenges for populating the KGs with data. The pragmatic approach we adopted of creating online spreadsheets for each group, based on a draft schema, enabled group members without any training in semantic technologies to populate their databases. However, it is yet to be determined whether enthusiasm and energy for adding and updating data will persist over time and how the spreadsheets can be synchronized with the RDF triple store. There has been limited work on the collaborative development of knowledge graphs, but we may be able to draw from experiences from a few large-scale examples, such as DBpedia [84], Wikidata [85], as well as metadata librarianship [86,87].

Manual approaches to KG population require considerable personnel effort. Natural language processing (NLP) provides another method with the potential to bring large amounts of data into a KG quickly. Wang and Stewart [88] demonstrated how NLP could be used to populate a graph on natural hazards, including spatiotemporal information, from web-based news reports. However, automated KG population can be quite difficult,

particularly as the diversity of subject matter and entity types increases [15]. Future research could investigate the potential for a web/document crawling system that could ingest a variety of sources from partners and linked websites.

A second challenge to ongoing collaborative KG development is schema management. We began data entry for the three case studies with a common data model, but somewhat different needs for each case quickly emerged. The semantic triple-store approach does not actually require any fixed schema (in contrast to more traditional tabular approaches), but if classes, properties, and relationships are not in sync, they cannot be queried across different data stores. Our attempts to synthesize the evolving schemas through ongoing communications between cases underlined the need for parsimonious approaches to do so. Longer term strategies with potential include continued alignment with other existing ontologies [89], using common, tested design patterns [90], and various automated approaches to ontology alignment/matching [91].

Much of the potential for the semantic data approach lies in the ability to link to external data sources. Hitzler (2021) describes some of the largest public data stores available, including Wikidata (>66 million data items) [92], Dbpedia (>6 million) [93], and the Google Knowledge Graph (>5 billion) [94]. Large geographic data stores include GeoNames (>7.5 million features) [95] and OpenStreetMap (>7 billion nodes) [96]. Major sources of biophysical ontologies and linked data include the ESIP Community Ontology Repository [97], Earthcube [98], and the Open Biological and Biomedical Ontology (OBO) Foundry [99]. Despite the common commitment to linked open data, each of these resources can be challenging to use because of the variety of ontologies, ontology structures, and access interfaces. Future research could involve identifying REDP needs that could be potentially met by such external sources.

6.4. Geospatial Aspects

Finally, we came to this work with a strong background in spatial analysis. Although the top need seen in the REPDs was networking people and organizations rather than spatial analysis, many of these questions still had spatial elements. At least in the field of environmental data, most geospatial tools and data stores are not directly available to semantic data tools and approaches. Geospatial data in semantic data have been quite limited. The most common spatial approach in linked data is simply providing a latitude/longitude location of an entity (“point of interest” or POI). POIs, while useful for many applications, have limited utility for our case study needs or describing work being done in the environment, which typically involves irregularly shaped areas of interest, whether human or naturally defined. The GeoSPARQL standard greatly enhances the ability to represent and manipulate spatial data in RDF. It provided methods to meet our users’ needs via the import of selected spatial data into the triple store, but this import process creates an additional workflow, a demand for specialized knowledge, and will be difficult to sustain if spatial data are maintained and updated in separate GIS systems. Some tools are now available for moving data in the opposite direction, from semantic stores into GIS environments. Mai et al. [100] developed such a framework and toolbox for bringing linked data from the web into the popular ArcGIS software. This toolbox is potentially quite useful to REDP participants with GIS skills, but it is unclear if it can address user queries that are nonspatial, and it also does not address the problem of keeping GIS and semantic data synchronized. Another approach receiving attention is pairing semantic technologies with industry standards for online GIS interoperability [101]. Both Zhang et al. [102] and Li et al. [103] demonstrate methods for using semantics to construct queries across heterogeneous data sources, which are then translated to the Web Feature Service (WFS) standard, executed in the online GIS environment, and results returned and recombined using the semantic specifications. While these are still prototypes, not readily adoptable by our REPD groups, the k.Lab software stack produced by the Integrated Modelling Partnership is a more mature product [104,105]. It includes a web browser geospatial interface linked to a semantic modeling framework. From the viewpoint of this project, its

current limitations are that all data must reside on or be linked from their central server and that it does not directly interface with RDF data sources (although they note that an RDF adaptor is in development).

7. Conclusions

Our work with three multi-stakeholder landscape collaboratives confirmed the results of an earlier broadscale survey that better tools for information sharing are among their top needs. Semantic technologies, such as linked data and knowledge graphs, have potential to increase the abilities of such collaboratives to share information from their different repositories as well as build new information infrastructures. We have provided a methodology for elucidating these needs in enough detail to design specific semantic data structures, and we built such a data structure for the greatest shared need, the ability to track inter-related people, projects, organizations, and what they produce (PPOp). We believe the PPOp framework we have created has wide applicability beyond REDPs to virtually any type of collaborative venture where work efforts are being explored, proposed, funded, implemented, evaluated, or shared.

While we believe the emerging PPOp schema could be useful to many groups going forward, the creation of an open knowledge network based on PPOp will require significant progress on a number of fronts. We found the semantic software tools available to be too complex and piecemeal to be usable by groups such as REDPs with limited technical capacity and funding. More development of simpler, browser-based front-ends is needed. Users will need easier ways to enter and maintain these data. Automating the extraction of such data from existing sources (websites, documents) using natural language processing could speed up this process. Finally, there are important spatial aspects to our PPOp entities. The GeoSPARQL standard provides a means to handle spatial data and queries in linked data formats (RDF/OWL); however, the maintenance and synchronization of both semantic and geospatial datastores is a serious challenge. Emerging hybrid approaches that dynamically mediate between linked data and online GIS standards show promise for addressing this issue.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijgi10120823/s1>. Table S1: Case Study Questions and Synthesis Clusters, Table S2: PPOp Program class properties.

Author Contributions: Conceptualization, Sean N. Gordon, Philip J. Murphy, Patrick Huber, Allan Hollander and John A. Gallo; Data curation, Philip J. Murphy and Allan Hollander; Formal analysis, Philip J. Murphy, Allan Hollander and Sean N. Gordon; Funding acquisition, Sean N. Gordon, Philip J. Murphy, John A. Gallo and Piotr Jankowski; Investigation, Philip J. Murphy, John A. Gallo and Ann Edwards; Methodology, Sean N. Gordon, Philip J. Murphy; Project administration, Sean N. Gordon; Software, Phillip Murphy and Allan Hollander; Supervision, Sean N. Gordon; Validation, Sean N. Gordon; Visualization, Philip J. Murphy; Writing—original draft, Sean N. Gordon, Philip J. Murphy, John A. Gallo, and Ann Edwards; Writing—review and editing, Patrick Huber, Allan Hollander, and Piotr Jankowski. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US National Science Foundation, Office of Integrative Activities, Program on Convergence Acceleration, project number 1937908.

Data Availability Statement: Manual KG approach data: data schemas (ontologies) are available in the public repositories cited in the text; data instances may be available at the discretion of the collaborating REDPs.

Acknowledgments: We wish to thank Naicong Li, Steve Paplanus and Yuanyuan Tian for help with development and testing of the PPOp ontology.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Scarlett, L.; McKinney, M. Connecting People and Places: The Emerging Role of Network Governance in Large Landscape Conservation. *Front. Ecol. Environ.* **2016**, *14*, 116–125. [CrossRef]
- Wondollock, J.M.; Yaffee, S.L. (Eds.) *Making Collaboration Work: Lessons from Innovation in Natural Resource Management*; Island Press: Washington, DC, USA, 2000; ISBN 1559634618.
- Lemos, M.C.; Agrawal, A. Environmental Governance. *Annu. Rev. Environ. Resour.* **2006**, *31*, 297–325. [CrossRef]
- Bixler, R.P.; McKinney, M.; Scarlett, L. Forging New Models of Natural Resource Governance. *Front. Ecol. Environ.* **2016**, *14*, 115. [CrossRef]
- Abrams, J. The Emergence of Network Governance in U.S. National Forest Administration: Causal Factors and Propositions for Future Research. *For. Policy Econ.* **2019**, *106*, 101977. [CrossRef]
- Forman, R.T.T. *Land Mosaics: The Ecology of Landscapes and Regions*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 1995; ISBN 978-0-521-47462-7.
- Johnson, K.N.; Swanson, F.; Herring, M.; Greene, S. (Eds.) *Bioregional Assessments: Science at the Crossroads of Management and Policy*; Island Press: Washington, DC, USA, 1999; ISBN 1559636572.
- Thomsen, J.M.; Caplow, S.C. Defining Success over Time for Large Landscape Conservation Organizations. *J. Environ. Plan. Manag.* **2017**, *60*, 1153–1172. [CrossRef]
- NLC. *Assessing the State of Landscape Conservation Initiatives in North America*; Network for Landscape Conservation: Bozeman, MT, USA, 2018. Available online: <https://landscapeconservation.org/our-work/landscape-conservation-initiatives-survey/> (accessed on 11 May 2021).
- USDOI; USDA; USDOC; CEQ. *Conserving and Restoring America the Beautiful*; US Department of Interior; US Department of Agriculture; US Department of Commerce; Council on Environmental Quality: Washington, DC, USA, 2021. Available online: <https://www.doi.gov/sites/doi.gov/files/report-conserving-and-restoring-america-the-beautiful-2021.pdf> (accessed on 15 June 2021).
- NLC. *Pathways Forward: Progress and Priorities in Landscape Conservation*; Network for Landscape Conservation: Bozeman, MT, USA, 2018. Available online: https://landscapeconservation.org/wp-content/uploads/2018/08/Pathways-Forward_2018_NLC.pdf (accessed on 11 May 2021).
- Campellone, R.M.; Chouinard, K.M.; Fisichelli, N.A.; Gallo, J.A.; Lujan, J.R.; McCormick, R.J.; Miewald, T.A.; Murry, B.A.; Pierce, D.J.; Shively, D.R. The ICASS Platform: Nine Principles for Landscape Conservation Design. *Landsc. Urban Plan.* **2018**, *176*, 64–74. [CrossRef]
- Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
- Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. *Sci. Am.* **2001**, *284*, 34–43. [CrossRef]
- Hogan, A.; Blomqvist, E.; Cochez, M.; d'Amato, C.; de Melo, G.; Gutierrez, C.; Gayo, J.E.L.; Kirrane, S.; Neumaier, S.; Polleres, A.; et al. Knowledge Graphs. 2020. Available online: <https://arxiv.org/abs/2003.02320> (accessed on 28 June 2021).
- Hitzler, P. A Review of the Semantic Web Field. *Commun. ACM* **2021**, *64*, 76–83. [CrossRef]
- SDSC. Spatial Decision Support Knowledge Portal. Available online: <http://sdsportal.sdsconsortium.org/> (accessed on 28 June 2021).
- Li, N.; Raskin, R.; Goodchild, M.; Janowicz, K. An Ontology-Driven Framework and Web Portal for Spatial Decision Support. *Trans. GIS* **2012**, *16*, 313–329. [CrossRef]
- State of Washington. *Puget Sound Partnership—Created*; 2007; Volume 90.71.210. Available online: <https://app.leg.wa.gov/rcw/default.aspx?cite=90.71.210> (accessed on 15 October 2021).
- Global Water Partnership. What Is IWRM? Available online: <https://www.gwp.org/en/GWP-CEE/about/why/what-is-iwrm/> (accessed on 2 July 2021).
- Puget Sound Partnership. Boards Overview. Available online: <https://www.psp.wa.gov/partnership-boards-overview.php> (accessed on 2 July 2021).
- Puget Sound Partnership. *The 2018–2022 Action Agenda for Puget Sound*; 2018. Available online: https://psp.wa.gov/action_agenda_center.php (accessed on 2 July 2021).
- PSP Science Panel. *Priority Science to Support Recovery of the Puget Sound Ecosystem: A Science Work Plan for 2020–2024*; Puget Sound Partnership: Olympia, WA, USA, 2020. Available online: <https://pspwa.app.box.com/s/e81y0ap941ntik8o0me8o1lo6v12act1> (accessed on 1 July 2021).
- NWIFC. *2020 State of Our Watersheds*; Northwest Indian Fisheries Commission: Olympia, WA, USA, 2020. Available online: <https://nwifc.org/publications/state-of-our-watersheds/> (accessed on 19 September 2019).
- CCG. *Puget Sound Partnership 2018–2022 Action Agenda After Action Review*; Cascadia Consulting Group: Seattle, WA, USA, 2020. Available online: <https://pspwa.app.box.com/s/71kh95t3vx5ltu8931nntd0jdvpqk41k> (accessed on 15 August 2021).
- Puget Sound Partnership. Puget Sound Vital Signs. Available online: <https://vitalsigns.pugetsoundinfo.wa.gov/> (accessed on 2 July 2021).
- PSEMP. *PSEMP Spatial Data Work Group RECOMMENDATION*; Puget Sound Partnership: Olympia, WA, USA, 2019. Available online: <https://pspwa.app.box.com/s/9lytxocq3lpbuhwd0sxz9rq44i00tvb> (accessed on 18 July 2021).

28. SBCRPP. Regional Priority Plan to Reduce Wildfire Risk and Improve Forest Health in Santa Barbara County. Available online: <https://www.sbcwildfireresilience.org/> (accessed on 2 July 2021).
29. CEC. Climate Resilience Roundtables. Available online: <https://resource.cecsb.org/climate-resilience-roundtables/> (accessed on 2 July 2021).
30. CRCD. Cachuma Resource Conservation District. Available online: <https://www.rcdsantabarbara.org/> (accessed on 2 July 2021).
31. CBI. Data Basin. Available online: <https://databasin.org/> (accessed on 2 July 2021).
32. Sheehan, T.; Gough, M. A Platform-Independent Fuzzy Logic Modeling Framework for Environmental Decision Support. *Ecol. Inform.* **2016**, *34*, 92–101. [CrossRef]
33. Miewald, T.; Evans-Peters, S.; Bierly, K. *Cascades to Coast Collaborative Conservation Targets, Key Ecological Attributes and Criteria for Spatial Design*; Cascades to Coast Collaborative: Portland, OR, USA, 2020.
34. PNCLC. *Working Lands and Conservation Community Meeting: Chehalis, Washington*; Pacific Northwest Coast Landscape Collaborative, 26 October 2018. Available online: <https://medium.com/wild-without-end/working-lands-and-conservation-4df09ed29b1a> (accessed on 25 November 2021).
35. PNCLC. *Working Lands and Conservation Community Meeting: Warrenton, Oregon*; Pacific Northwest Coast Landscape Collaborative, 15 April 2019. Available online: file:///C:/Users/MDPI/AppData/Local/Temp/Clatsop_County_Long_Range_Plan_2019-1.pdf (accessed on 25 November 2021).
36. Gallo, J.; Aplet, G.; Greene, R.; Thomson, J.; Lombard, A. A Transparent and Intuitive Modeling Framework and Software for Efficient Land Allocation. *Land* **2020**, *9*, 444. [CrossRef]
37. W3C. RDF–Semantic Web Standards. Available online: <https://www.w3.org/RDF/> (accessed on 2 July 2021).
38. W3C. OWL 2 Web Ontology Language Document Overview (Second Edition). Available online: <https://www.w3.org/TR/owl2-overview/> (accessed on 2 July 2021).
39. W3C. SPARQL 1.1 Query Language. Available online: <https://www.w3.org/TR/sparql11-query/> (accessed on 2 July 2021).
40. Graves, M.; Constabaris, A.; Brickley, D. FOAF: Connecting People on the Semantic Web. *Cat. Classif. Q.* **2007**, *43*, 191–202. [CrossRef]
41. Mitchell, S.; Conlon, M. *VIVO: Ontology for Researcher Discovery (Version 1.7)*; 2018. Available online: <https://bioportal.bioontology.org/ontologies/VIVO/?p=summary> (accessed on 2 July 2021).
42. W3C. The Organization Ontology. Available online: <https://www.w3.org/TR/vocab-org/> (accessed on 2 July 2021).
43. Schema.Org. Available online: <https://schema.org/> (accessed on 5 December 2020).
44. Hollander, A.D.; Geith, C.; Lange, M.C. An Expertise Ontology for Cooperative Extension. In *Proceedings of the CEUR Workshop Proceedings*; Sun SITE Central Europe: Aachen, Germany, 2018. Available online: http://ceur-ws.org/Vol-2285/ICBO_2018_paper_30.pdf (accessed on 12 July 2019).
45. Hollander, A. *PPOD Ontology*; 2019. Available online: <https://github.com/adhollander/ppod> (accessed on 13 June 2021).
46. Hollander, A.D.; Hoy, C.; Huber, P.R.; Hyder, A.; Lange, M.C.; Latham, A.; Quinn, J.F.; Riggle, C.M.; Tomich, T.P. Toward Smart Foodsheds: Using Stakeholder Engagement to Improve Informatics Frameworks for Regional Food Systems. *Ann. Am. Assoc. Geogr.* **2020**, *110*, 535–546. [CrossRef]
47. Global Forum on Agricultural Research. AgriVIVO Ontology. Available online: <http://agrivivoes.com/page/technical-details> (accessed on 13 June 2021).
48. Mungall, C. Relation Ontology. Available online: <http://www.obofoundry.org/ontology/ro.html> (accessed on 13 June 2021).
49. Huber, P.R.; Springer, N.P.; Hollander, A.D.; Haden, V.R.; Brodt, S.; Tomich, T.P.; Quinn, J.F. Indicators of Global Sustainable Sourcing as a Set Covering Problem: An Integrated Approach to Sustainability. *Ecosyst. Health Sustain.* **2015**, *1*, 1–8. [CrossRef]
50. Schneider, A.; Ingram, H. Behavioral Assumptions of Policy Tools. *J. Politics* **1990**, *52*, 510–529. [CrossRef]
51. Bibliographic Ontology (BIBO). Available online: <https://www.dublincore.org/specifications/bibo/bibo/> (accessed on 5 December 2020).
52. Yan, J.; Lv, T.; Yu, Y. Construction and Recommendation of a Water Affair Knowledge Graph. *Sustainability* **2018**, *10*, 3429. [CrossRef]
53. Senderov, V.; Simov, K.; Franz, N.; Stoev, P.; Catapano, T.; Agosti, D.; Sautter, G.; Morris, R.A.; Penev, L. OpenBiodiv-O: Ontology of the OpenBiodiv Knowledge Management System. *J. Biomed. Semant.* **2018**, *9*, 5. [CrossRef] [PubMed]
54. Kalabokidis, K.; Athanasios, N.; Vaitis, M. OntoFire: An Ontology-Based Geo-Portal for Wildfires. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 3157–3170. [CrossRef]
55. W3C Semantic Web Interest Group. Basic Geo (WGS84 Lat/Long) Vocabulary. Available online: <https://www.w3.org/2003/01/geo/> (accessed on 5 December 2020).
56. Koubarakis, M.; Kyzirakos, K. Modeling and Querying Metadata in the Semantic Sensor Web: The Model StRDF and the Query Language StSPARQL. In *Proceedings of the Semantic Web: Research and Applications*; Aroyo, L., Antoniou, G., Hyvönen, E., ten Teije, A., Stuckenschmidt, H., Cabral, L., Tudorache, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 425–439. [CrossRef]
57. OGC. *GeoSPARQL—A Geographic Query Language for RDF Data*; Open Geospatial Consortium: Arlington, VA, USA, 2012. Available online: <https://www.ogc.org/standards/geosparql> (accessed on 26 November 2021).
58. OGC GeoSPARQL. Wikipedia 2021. Available online: https://en.wikipedia.org/w/index.php?title=OGC_GeoSPARQL&oldid=1049804887 (accessed on 25 November 2021).

59. Tudorache, T.; Nyulas, C.; Noy, N.F.; Musen, M.A. WebProtege: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web. *Semant. Web* **2013**, *4*, 89–99. [CrossRef] [PubMed]
60. WebProtege–Protege Wiki. Available online: <https://protegewiki.stanford.edu/wiki/WebProtege> (accessed on 25 November 2021).
61. Gruff | AllegroGraph 7.2.0. Available online: <https://franz.com/agraph/support/documentation/current/gruff.html> (accessed on 25 November 2021).
62. Semantic MediaWiki. Available online: https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki (accessed on 25 November 2021).
63. Wbd (MapServer). Available online: <https://hydro.nationalmap.gov/arcgis/rest/services/wbd/MapServer> (accessed on 25 November 2021).
64. Hollander, A. PPODtottl. 2021. Available online: <https://github.com/adhollander/PPODtottl> (accessed on 25 November 2021).
65. RDFLib. 2021. Available online: <https://github.com/RDFLib/rdfib> (accessed on 25 November 2021).
66. Musen, M.A.; Protégé Team. The Protégé Project: A Look Back and a Look Forward. *AI Matters* **2015**, *1*, 4–12. [CrossRef]
67. Protégé. Available online: <https://protege.stanford.edu/> (accessed on 25 November 2021).
68. AllegroGraph. Available online: <https://allegrograph.com/> (accessed on 25 November 2021).
69. Smythe, K.D.; Bernabo, J.C.; Carter, T.B.; Jutro, P.R. Focusing Biodiversity Research on the Needs of Decision Makers. *Environ. Manag.* **1996**, *20*, 865–872. [CrossRef]
70. Steiner Davis, M.L.E.; Tenopir, C.; Allard, S.; Frame, M.T. Facilitating Access to Biodiversity Information: A Survey of Users' Needs and Practices. *Environ. Manag.* **2014**, *53*, 690–701. [CrossRef]
71. Ryan, C.M.; Cervený, L.K. Wildland Fire Science for Management: Federal Fire Manager Information Needs, Sources, and Uses. *West. J. Appl. For.* **2011**, *26*, 126–132. [CrossRef]
72. Gober, P.; Larson, K.L.; Quay, R.; Polsky, C.; Chang, H.; Shandas, V. Why Land Planners and Water Managers Don't Talk to One Another and Why They Should! *Soc. Nat. Resour.* **2013**, *26*, 356–364. [CrossRef]
73. Rayner, S.; Lach, D.; Ingram, H. Weather Forecasts Are for Wimps: Why Water Resource Managers Do Not Use Climate Forecasts. *Clim. Chang.* **2005**, *69*, 197–227. [CrossRef]
74. Guston, D.H. Boundary Organizations in Environmental Policy and Science: An Introduction. *Sci. Technol. Hum. Values* **2001**, *26*, 399–408. [CrossRef]
75. Ahearn, S.C.; Icke, I.; Datta, R.; DeMers, M.N.; Plewe, B.; Skupin, A. Re-Engineering the GIS&T Body of Knowledge. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 2227–2245. [CrossRef]
76. Shotton, D.; Peroni, S. FRAPO: The Funding, Research Administration and Projects Ontology. 2017. Available online: <https://sparontologies.github.io/frapo/current/frapo.html> (accessed on 2 July 2021).
77. Alatrish, E. Comparison Some of Ontology. *J. Manag. Inf. Syst.* **2013**, *8*, 018–024.
78. Clunis, J.S. Comparative Survey of Ontology Editors for the Semantic Web. In *iConference 2019 Proceedings*; iSchools: Grandville, MI, USA, 15 March 2019. [CrossRef]
79. KgBase–No-Code Knowledge Graphs. Available online: <https://www.kgbase.com/> (accessed on 25 November 2021).
80. Gra.Fo. Available online: <https://gra.fo/> (accessed on 25 November 2021).
81. Chaudhri, V.K. Stanford CS 520 Knowledge Graphs. Available online: <https://web.stanford.edu/class/cs520/> (accessed on 28 August 2021).
82. Sack, H.; Alam, M. Knowledge Graphs. 2020. Available online: <https://open.hpi.de/courses/knowledgegraphs2020> (accessed on 19 July 2019).
83. Open Textbooks. Available online: <https://www.oercommons.org/hubs/open-textbooks> (accessed on 25 November 2021).
84. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. DBpedia: A Nucleus for a Web of Open Data. In *Lecture Notes in Computer Science*; Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 722–735. ISBN 978-3-540-76298-0.
85. Vrandečić, D.; Krötzsch, M. Wikidata: A Free Collaborative Knowledgebase. *Commun. ACM* **2014**, *57*, 78–85. [CrossRef]
86. Schreier, P.E. Sinopia: A New Linked-Data Editing Environment Designed for Libraries. In *Proceedings of the Metadata and Semantic Research*; Garoufallou, E., Fallucchi, F., William De Luca, E., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 425–430. [CrossRef]
87. Thalhath, N.; Nagamori, M.; Sakaguchi, T.; Sugimoto, S. Wikidata Centric Vocabularies and URIs for Linking Data in Semantic Web Driven Digital Curation. In *Proceedings of the Metadata and Semantic Research*; Garoufallou, E., Ovalle-Perandones, M.-A., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 336–344. [CrossRef]
88. Wang, W.; Stewart, K. Spatiotemporal and Semantic Information Extraction from Web News Reports about Natural Hazards. *Comput. Environ. Urban Syst* **2015**, *50*, 30–40. [CrossRef]
89. Fonseca, A.M.; Camolesi, L. Refactoring Rules for Graph Databases. In *New Contributions in Information Systems and Technologies*; Rocha, A., Correia, A.M., Costanzo, S., Reis, L.P., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2015; Volume 353, pp. 33–42. ISBN 978-3-319-16485-4.
90. Shimizu, C.; Hirt, Q.; Hitzler, P. MODL: A Modular Ontology Design Library. *CoRR* **2019**. Available online: <http://arxiv.org/abs/1904.05405> (accessed on 18 September 2020).
91. Noy, N.F. Semantic Integration: A Survey of Ontology-Based Approaches. *SIGMOD Rec.* **2004**, *33*, 65–70. [CrossRef]
92. Wikidata. Available online: https://www.wikidata.org/wiki/Wikidata:Main_Page (accessed on 25 November 2021).

93. DBpedia. Available online: <https://www.dbpedia.org/> (accessed on 25 November 2021).
94. Google Knowledge Graph Search API. Available online: <https://developers.google.com/knowledge-graph> (accessed on 25 November 2021).
95. GeoNames Search Webservice. Available online: <https://www.geonames.org/export/geonames-search.html> (accessed on 25 November 2021).
96. OSM Sophox Service. Available online: <https://sophox.org/> (accessed on 25 November 2021).
97. ESIP Federation Community Ontology Repository Services. Available online: <http://cor.esipfed.org/> (accessed on 25 November 2021).
98. GeoCODES. Available online: <https://geocodes.earthcube.org/> (accessed on 25 November 2021).
99. The Open Biological and Biomedical Ontology (OBO) Foundry. Available online: <http://obofoundry.org/> (accessed on 25 November 2021).
100. Mai, G.; Janowicz, K.; Yan, B.; Scheider, S. Deeply Integrating Linked Data with Geographic Information Systems. *Trans. GIS* **2019**, *23*, 579–600. [[CrossRef](#)]
101. OGC Standards. Available online: <https://www.ogc.org/docs/is> (accessed on 25 November 2021).
102. Zhang, C. Ontology for Geospatial Semantic Interoperability. In *The Geographic Information Science & Technology Body of Knowledge*; Wilson, J.P., Ed.; University Consortium for GIS Science: Washington, DC, USA, 2019. [[CrossRef](#)]
103. Li, W.; Song, M.; Tian, Y. An Ontology-Driven Cyberinfrastructure for Intelligent Spatiotemporal Question Answering and Open Knowledge Discovery. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 496. [[CrossRef](#)]
104. Villa, F.; Bagstad, K.J.; Voigt, B.; Johnson, G.W.; Portela, R.; Honzák, M.; Batker, D. A Methodology for Adaptable and Robust Ecosystem Services Assessment. *PLOS ONE* **2014**, *9*, e91001. [[CrossRef](#)]
105. Integratedmodelling.Org | Resources for a Semantic Modeling Community. Available online: <https://integratedmodelling.org/> (accessed on 25 November 2021).