Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace

Dominic Oldman(✉) and Diana Tanase(✉)

British Museum, London, UK
{doldman,dtanase}@britishmuseum.org

Abstract. ResearchSpace is an open source platform designed at the British Museum to help establish a community of researchers, where their underlying activities are framed by *data sharing*, *active engagement in formal arguments*, and *semantic publishing*. Using Semantic Web languages and technologies, the innovations of the system are shaped by a social conceptualisation of the graph-based representation of information. This is employed by integrated semantic components aimed at subject experts that offer mechanisms to create, annotate, assert, argue, search, cite, and justify data-driven research. This paper showcases a new onto-epistemological approach that supports researchers to contribute to a growing and sustainable corpus of knowledge that has history, not just provenance, built-in. It describes our considerations in designing for interdisciplinary collaboration, usability and trust in the digital space, highlighted by use cases in archaeology, art history, and history of science.

1 Introduction

The ResearchSpace open source project[1] is a vision of the Andrew W. Mellon Foundation, a major funder of humanities and cultural heritage research. This article describes the system developed based on this vision. Designed at the British Museum and developed in partnership with Metaphacts[2], Semantic Web technologies are combined to enable creation, sharing and linking of information in different configurations and levels of de-centralisation, between multiple confederated instances. The emphasis is on a representation of information that integrates, preserves multiple perspectives, and promotes collaborative research. The project's long term goal is to build a community of researchers that open and share their *data*, *knowledge*, *research practices*, and *arguments* with each other.

Museums, libraries, archives and other research institutions are often referred to as *knowledge* or *memory* institutions, but their contribution to the Web of

[1] http://github.com/researchspace/researchspace.
[2] https://metaphacts.com.

D. Vrandečić et al. (Eds.): ISWC 2018, LNCS 11137, pp. 325–340, 2018.
https://doi.org/10.1007/978-3-030-00668-6_20

Data is often disconnected from subject experts. The system presented here attempts to reconnect with those experts and focuses on knowledge representation of data for research, rather than just mapping traditional forms of digital documentation to Linked Data. The latter records substantive information about material things mainly for referential purposes (e.g. inventories, catalogues). This data, not originally designed for wider reuse, has become the basis for online publication and aggregation, but only represents a small part of overall institutional knowledge. Some types of administrative public data provide significant informational value particularly when transferred to formats like Resource Description Framework (RDF) [1]. The situation is less clear for complex subject areas that have wide historical significance with various and often contentious perspectives, and where new knowledge is constantly being identified. This is the case for art, cultural and historical knowledge, which is part of an ever changing understanding of the world and its development.

Considering these domain constraints, ResearchSpace was developed with an onto-epistemological approach to information representations aimed at research, relations discovery and dissemination using suitably expressive structures presented in an accessible form (see Sect. 2 for existing approaches). The current software implementation allows for:

- Assertion and argumentation models for tracing the multiple perspectives on history and the material world.
- Creation of semantic data and narratives, as well as expert-led refinements and expansions of existing data through a growing graph structure, namely the *knowledge graph*, that formally captures different worldviews and their provenance.
- Multi-level visual representation of resources (e.g. historical processes and entities), structured and comprehensive exploration of resources based on maps, timelines, charts, comparative image overlays, search and browsing across different heterogeneous datasets.
- Presentations of findings that explicitly record and describe researchers' views expressed in a graph that connects narrative, data, processes, and arguments.

We argue that such systems must be approached as a *psychological tool*[3], a mediation space, in which, despite recent trends in computing, and in particular machine learning, the focus should remain on the human as a builder of knowledge by reshaping the knowledge graph. This addresses the issue of how the computer deals with change and how thinking that we normally associate with writing narratives can be transferred and evolved within a digital environment. To this aim we propose a broader definition of the knowledge graph that encapsulates the unique nature of the relationship between humans and computers

[3] Lev Vygotsky used *psychological tools* in the context of sociocultural theory of cognitive development. They are described as "the form of mediation needed for the emergence of conceptualised thinking and the tools in question include, in addition to speaking and writing, gestures, sign systems such as maps and diagrams, and mnemonic techniques" [5].

necessary for a knowledge system to operate effectively (see Sect. 3 for a description of the system). In doing so, we put forward the following considerations for digital research systems:

- Computer systems should be designed to actively encourage human knowledge production.
- Knowledge is cumulative and should be built upon to reduce fragmentation.
- Research increases in effectiveness as a social and collaborative activity embracing formal models of argumentation to investigate contradictions.
- Computer systems that aim to reflect aspects of the world need to deal with changing knowledge and support questioning established history.
- Data should be contextualised to reflect complexity and presented in accessible ways to wider audiences, not just experts.

These key points constitute guiding principles for existing projects that use ResearchSpace in archaeology, art history, and history of science (see Sect. 4).

2 Challenges for a Digital Research Space

Throughout this project's development we identified four groups of challenges relating to data, infrastructure, publishing and the researchers' mindset. In this section we contrast our approach with existing work.

2.1 Creating, Integrating, Sharing and Reuse of Data

In the last decade several museums including the British Museum have opened their collection data to the World Wide Web. Yet, this is problematic since cultural heritage collection data systems were designed for internal administration by specialist users, where the shortfall in data specificity, ambiguities, or uncertainties are compensated by the knowledge of expert users who interpret it. The language and the knowledge required to understand the original meaning behind the data is not accessible to external users when this data is openly published in Linked Data format. It is here that we see the tension between those working with Linked Data focusing on reuse, and the needs of domain experts. Best practice[4] notes on data publication make no explicit recommendations based on source knowledge characteristics or consider the quality of the mapping process to Linked Data in terms of its usefulness for target applications.

We advocate for ontological representation of data from the cultural heritage domain, which provides a framework for a high degree of semantic meaning and contextual structure to be expressed, but also the ability to create inferred presentations for different audiences that communicate appropriately using the same underlying data. Therefore, the challenge goes beyond solving the technical tasks of creating, sharing, and reuse of data, but rather to consider its wider long term use and purpose.

[4] https://www.w3.org/TR/dwbp/.

2.2 An Infrastructure for Transferability of Knowledge

Just as data design can be affected by top down decisions, the same applies to infrastructure and this leads to concerns about the nature of a digital infrastructure and how this relates to its users. A common starting point is the National Science Foundation's Atkins report (2003) [2] which provided a general description as part of a case for investment in new physical infrastructures. It gave an overview of the technical components (hardware and software) and the multidisciplinary expertise required to operate it. Other studies, for example [19], emphasized that the components should be a conduit for interactions between technology and people, with the *researcher an integral component of the application execution*. The need for visualisation interfaces for human interaction was advocated as a crucial aspect.

In [17], a further demand is articulated from a research infrastructure, namely the ability to link to real research outcomes and prevent technology from distorting them. Hence, the importance of transparency at the level of data and process. A Digital Humanities model proposed by [9] sought to embed scholarly activities into the definition of infrastructure. These were refined by the Scholarly Domain Model (SDM) based on an initial, now expanded, set of scholarly primitives, namely discovering, annotating, comparing, referring, sampling, illustrating, and representing. Also [14,15] argued that a lack of attention to data design and data outputs compared to function, risked making research systems ineffectual by not having the required semantic elements for reliable knowledge building, also discussed by [3].

This literature review charts the continuous refinement of what is considered a good infrastructure for digital research projects. However, the value of having such an infrastructure in the cultural heritage and arts domain is to "explore theories, ways of perceiving, ways of knowing; to enter into other mindsets and world-views" [20]. This requires interpretative work that is not possible just with technologies, but needs human experts. The Semantic Web and Linked Data solutions address some types of data integration problem, but ignore the underlying need of experts to collaboratively grow information over time in a relevant way for their research. ResearchSpace makes this possible by using Semantic Web technologies in applications designed specifically for knowledge workers. The system presents the user with an interface to make research activities like creation, discovery, enrichment, argument formulation and publication - intuitive, while in the background it employs rich ontologies to record interactions as RDF. The new triples trace the connections between research activities thus adding the missing historical dimension.

2.3 Semantic Scholarly Publishing

The combination of narrative and data in semantic publishing is of increasing interest to the Semantic Web community. In [11], the authors provide a critique of current systems by pointing out that these solutions transform existing publications, rather than support direct semantic creation and publication. True semantic publishing should come directly from authors, and if possible as part of an integral knowledge generation activity not mediated by technologists which inevitably stifles the ability to use a digital environment as a place where modes of thinking actually take place.

Centralisation can lead to issues of data ownership, accessibility, bias and limitation of freedom of expression. Decentralisation of semantic publications and the existence of accessible tools that support it addresses these problems, but at the same time takes on aspects of traditional narratives and disciplinary fragmentation. These narratives host descriptive and analytical modes of communication. In descriptive narrative, such as those provided by Wikipedia, articles are restricted to *facts* about things (material and immaterial), events or people. The Wikipedia community is used to enforce a neutrality rather than express the different perspectives represented within it. Equally, traditionally authored analytical narratives use evaluation and comparison, inevitably selecting evidence that supports a particular hypothesis.

Decentralised semantic publication also draws upon this tradition using selected semantic elements that support certain perspectives without adequate comparative semantics to allow effective resolution of contradiction. Typically academic disciplines are silos of information and knowledge using their own narrative conventions that make interdisciplinary studies difficult. Narratives are highly heterogeneous, and have both linguistic complexity and ambiguity, but they have overlapping concepts, ideas and information that are important to building a history of interdisciplinary knowledge.

The descriptive narratives of Wikipedia have much in common with aggregated data services in that they tend towards an almost fixed structure and form. Aggregated data services, in attempting to deal with variable local resources, will force data to conform to a central model. In pursuit of being open to a range of audiences a dominant common denominator approach is developed based on a perception that general audiences require a reduction of complexity for reuse. These centralised services produce descriptive models with limited contextual richness and therefore narrow reuse value. These constraints are also apparent in cultural aggregations like Europeana, resulting in some researchers attempting to enrich the data using narratives, for example [13].

At the other extreme, completely decentralised publications also present problems for reuse and the ability to use the knowledge of different communities for progressive digital publication. Creating semantically enhanced publications, whether born semantic or whether semantically enriched existing publications, can use data to describe things difficult to describe precisely just through text. Equally, narratives can clarify and make accessible abstract data. Using structured data within narratives can potentially improve clarity and discovery, therefore support better assessments of a particular subject area. It also holds the possibility for computer inferences across publications. Decentralised articles which employ semantic enrichment, but embed raw data based on many different ontologies, place some limitations on both discovery and inference. A coherent framework of semantics used consistently across data and narratives allows greater scope to improve meaning within a particular narrative, but also across many narratives.

Applications like Dokieli [4] provide decentralised publishing environments in which the user can generate data enriched narratives and establish personal networks. However, reliance on completely decentralised modes of thinking risk continuing disciplinary fragmentation due to the lack of semantic integration and sometimes appropriateness (in terms of data representation) making analysis across genuine semantic publications no less problematic than with traditional publications.

2.4 Mindset and the Interpretation of the Knowledge Graph

While decentralised semantic publishing is still in its infancy, the whole notion of the Semantic Web remains opaque to many people. Many users have a narrow view of the role information systems have in their work, based on a dominant mindset around traditional database systems. Most institutions adopt information systems that fulfil an administrative, reference and operational mindset, rather than one which seeks to promote knowledge-based activity. There is a separation between what the computer provides and the intellectual processes and activities retained by the human operator, and used elsewhere. The aim of representing knowledge using the Semantic Web means that computer scientists and domain experts need to re-evaluate their relationship with these new types of information (knowledge base) system. This is required for a transition from traditional and individualistic methods of research, to collaborative and open research practices.

Domain experts need to be active in the design of the Semantic Web applications rather than just inform basic requirements gathering. In the same way that technologists cannot represent the knowledge of other domains, they are unable to implement appropriate knowledge systems without the direct involvement of source experts at the design level.

The situation is not helped by current technical definitions of a knowledge graph, not just because of the language used but also because they tend to be technology centered and fail to encapsulate the intellectual contribution of the user. The notion of knowledge graph has been discussed in [16] and further debated in [8]. One general definition views knowledge graphs as large networks of entities, their semantic types, properties, and relationships between them based on automatically derived and interlinked factual information from knowledge bases such as DBpedia[5] and others.

We propose a definition intended to influence the design of a knowledge oriented information system that recognises the wider role of the user in the creation of the graph-based representational structure at its core: *a knowledge graph is a continually changing informational structure that mediates between a human, the world and a computer. The graph itself is ontologically based and enhanced by human epistemology. These are closely linked in that the ontology provides*

[5] https://wiki.dbpedia.org/.

real world references and a structure of interrelated entities or processes, while the epistemology uses the graph to interpret and generate new knowledge. Growing the graph is based on both automated reasoning and crucially, collaborative human thinking and creativity.

This new conception of the knowledge graph, helps capture the nature of change in knowledge in a representation suitable for interdisciplinary scholarship enabling the ongoing process of bringing knowledge into being with historical provenance.

3 Connecting Researchers, Data and Practices

Lowestoft is a town on the most easterly point of Great Britain. It was featured in the 2015 British General Election as a coastal town economically depressed and mostly forgotten, but still hopeful and looking optimistically forward into the future. When we read the Wikipedia descriptive entry it is hard to understand where this optimism might come from, but when delving deeper into cultural heritage, and related resources, one finds a completely different impression from the encyclopaedic perspective. To the computer system a place is a static 'entity', but in our wider version of the knowledge graph definition, it is a changing 'process' with large numbers of relations to other processes, across time and space. The demonstration instance of the ResearchSpace system[6] uses an RDF graph based on a subset of the British Museum's catalogue data. This provides an example of semantic enhancement of an existing data resource achieved by mapping institutional data to an event-based ontology called CIDOC CRM (Conceptual Reference Model) [6]. ResearchSpace is setup to offer multiple paths of exploration and analysis. In a shared environment multiple researchers can investigate the history of Lowestoft afresh, or from a set of established resources (and perspectives) in a shared clipboard. This can contain predefined sets of semantic resources such as, charts, diagrams, arguments, and searches. Similarly, a Semantic Narrative can include the same resources taken from the clipboard and juxtaposed with text. Therefore, an existing narrative on Lowestoft might already hold defined searches or other resources that are relevant and act as an existing research object to build upon, or argue with. Typically, as a researcher finds relevant resources they are saved into a clipboard through a drag and drop mechanism and can be subsequently organised into different sets. Through the British Museum's data, a picture of Lowestoft's historical periods can be developed and knowledge from other sources added.

If a researcher needs to challenge existing British Museum data or assert completely new information then this can be added using the ontology rather than textual annotations. During this process of discovery and enrichment, the researcher will realise that Lowestoft has a long and rich history, and that it attracted many people, for example artists like, J.W.M. Turner, Muirhead Bone and Samuel Varsey. It produced fine porcelain exported around the world. Its position on the North Sea means that it has military associations, as well as

[6] https://demo.researchspace.org.

social and economic connections with other countries, particularly Scandinavia and Northern Europe. The Battle of Lowestoft in 1665, is documented within UK and Dutch museums. The military legacy and boat building tradition is reflected in objects in several cultural heritage institutions, including archival photographs, medals, newspaper cuttings, letters and so on. The ResearchSpace system is able to bring this data together semantically intact, and visualise it in multiple ways. The data found from direct connections to Lowestoft can have secondary semantic references to other parts of England, or the World, and the network of people with connections to Lowestoft becomes wider, including writers such as Joseph Conrad, Rudyard Kipling and Charles Dickens.

3.1 The Technology Stack

The ResearchSpace technology stack (see Fig. 1) builds on the metaphactory knowledge graph platform enabling customisation and extensibility of the interaction with the graph database (Blazegraph[7]) through the use of familiar open standards such as RDF and SPARQL, expressive ontologies for schema modeling based on CIDOC CRM[8], rules, constraints, and query specifications based on SPIN[9], W3C Web Components[10], W3C Open Annotation Data Model[11], and W3C Linked Data Platform Containers[12]. The platform is open source, integrating external tools including OntoDia[13], MIRADOR Image Viewer[14] with an IIIF Image Server[15]. Instantiating ResearchSpace for application projects involves creating templates, which are a mixture of HTML5[16], React Components[17] and Handlebars[18]. The custom HTML5/REACT Web components described in the next section are informed by domain experts, which operate on the result of SPARQL queries. They represent a selection of the ResearchSpace key features.

3.2 Semantic Components

Semantic Component: Knowledge Patterns They are predefined graph paths that express data creation, modification, deletion and visualisation. The use of the term *knowledge pattern* acknowledges their association with experts' needs for capturing, at various levels of detail, the contexts of processes involved in research. Technically, a knowledge pattern includes defining a set of SPARQL

7 https://www.blazegraph.com.
8 http://www.cidoc-crm.org.
9 http://spinrdf.org.
10 https://www.w3.org/TR/components-intro.
11 https://www.w3.org/TR/annotation-model.
12 https://www.w3.org/TR/ldp.
13 http://www.ontodia.org.
14 http://projectmirador.org.
15 http://iiif.io.
16 https://www.w3.org/TR/2010/WD-html5-20100624.
17 https://reactjs.org.
18 https://handlebarsjs.com.

Fig. 1. ResearchSpace platform architecture

1.1 statements but additional metadata is used for their integration into other components such as data input, visualisations, arguments and custom search systems. A single knowledge pattern combines with others but is transferable as an LDP Resource between different instances of ResearchSpace. Knowledge patterns are defined on an ongoing basis and subject experts are encouraged to learn the main system ontology which is essential to the ResearchSpace design. CIDOC CRM is a rich ontology including a growing number of specialisations with adoption in cultural heritage and beyond. It provides a contextual framework under which diverse and variable information can be integrated without homogenisation. Carefully designed UI exposes the ontology (if requested) to non-technical users encouraging involvement in the design of new knowledge patterns covering different areas of interest. For example, the project, Late Hokusai (see below), models patterns that describe the condition states of a woodblock over time, and its relationship with impressions (prints) derived from it. They help answer specific art history inquiries, but also address wider societal questions.

Semantic Component: Assertions and Arguments Formalisation of argumentation is a process of reasoning in support of an observation, idea, action, theory, or interpretation of facts, and is increasingly implemented by systems that support decision-making through social interactions [18]. The implementation of argumentation in ResearchSpace enables the creation of structured data assertions and arguments. These challenge existing entities and relations or make new assertions based on direct observation, the adoption of belief from others, or inference based on premises that are resources in the system (data, images, narratives, other assertions or arguments, etc.). These assertions and arguments have clear knowledge provenance. Arguments are resources that can be combined with other information resources in the system including explanatory narratives. ResearchSpace does not limit a particular argumentation logic but uses a model of human argumentation in which, "reasoning may not only consist of falsification or verification, but more generally of strengthening or weakening hypotheses, and a way to connect this model to an ontology of the domain of discourse" [7]. In this instance, the system uses the CIDOC CRM specialisation, CRMInf[19], to record the details of premises and conclusions. In Fig. 2, we visualized a very simple argument that a water colour representation of a boat named *Hope*, in the Lowestoft shipyard, was influenced by the works of Cornelius Varley. The expert's notes about the painting mention a connection to the British painter without recording the facts supporting this observation. What is essential for researchers is the ability to see the evolution, composition, and revision of arguments making explicit both the processes of argument-making and the states of belief at particular points in time in a composite inference [7]. This is relevant to the needs of a trusted digital environment with an embedded history of arguments. Using these principles, highly complex arguments can be constructed.

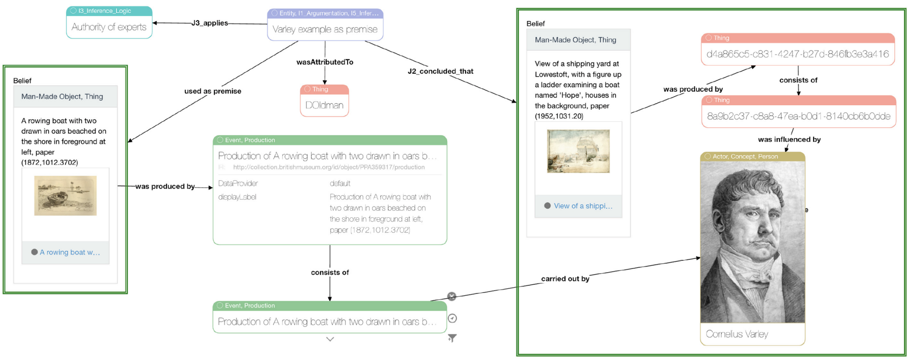

Fig. 2. Argument representation for expressing how Cornelius Varley influenced the production of a painting

Semantic Component: Data Enriched Narrative In the challenges section we outlined the problems with both centralised and decentralised publication. Centralisation imposes constraints on subject experts who should be directly responsible for their content. ResearchSpace provides the opportunity for semantic authoring, but within a particular community that agrees on an ontological framework. The objective is to ensure comparability, integration and a common purpose of building knowledge through different perspectives. These will almost certainly contain contradictions. Fully decentralised semantic publishing, even with data enrichment, has the same disadvantage as traditional analytical

[19] http://www.ics.forth.gr/isl/CRMext/CRMinf/docs/CRMinf-0.7.pdf.

narratives, in that contradictions and arguments are hard to identify, let alone resolve. In centralised systems contradictions create difficulties and are often avoided. The ability to integrate argument within the medium of narrative helps resolve analytical differences head on by fully understanding the nature of contradictions and providing an appreciation for the range of influences that cause change, and which individual researchers are unlikely to identify themselves. A third party non-technical visual editor, called Ory, was integrated to organise and visualise different RDF representations of content including ResearchSpace semantic resources into a narrative[20]. From the ResearchSpace clipboard, different resources are dragged into formatted blocks. For example, a Semantic Search can be saved as a resource, placed in the clipboard, and then dragged into a narrative. The Semantic Narrative component allows researchers to write text and use citations of contextualised data they, or others, have added, and visualise it appropriately. This means that embedded data can also be inspected by other readers. A particularly important example of this is the use of assertion and arguments providing the ability to combine the accessibility of narrative alongside the complex data it seeks to explain and to allow greater integration of information. It is important for the system to track and alert authors to new data, and particularly to arguments. They then can review and update their narratives, but also add and enrich entities to reflect new positions. The narrative acts as an accessible data reference point for performing these dynamic tasks that address the type of changes that computers are unable to interpret and respond to.

Semantic Component: Search Search is a classic task where the researcher formulates a question and the system answers with a set of relevant resources. ResearchSpace supports three different types of search scenarios. They differ in the way the system handles the formulation and transformation of a query into a set of resources. These are: (a) knowledge-graph driven search, (b) knowledge pattern-based search, and (c) text-based search.

Knowledge Graph Driven Search. Researchers expect to ask who, what, where, and when type of questions from computer information systems. ResearchSpace enables adding data and paths that support the formulation of ‘why’ questions based on arguments and semantic narratives. Using data from the British Museum, we built a semantic network of RDF triples that provides depth and detail to each entry. The solution to designing a UI based on a rich ontology, but with suitable recall and precision, uses an approach developed by the Foundation for Research and Technology Hellas (FORTH) [7]. The complexity of searching across a graph flows from using CIDOC CRM to capture sufficiently rich semantics. Exploring graph paths using exhaustive automatic reasoning would not be possible in real time. Therefore, a model for querying is employed based on six fundamental categories (FCs) that abstract the main entities in the graph: Thing (material and immaterial), Actor, Place, Time, Event, Concept. Relationships (properties) are similarly inferred creating a matrix of semantic shortcuts carving paths through the detailed informational space.

[20] https://github.com/ory.

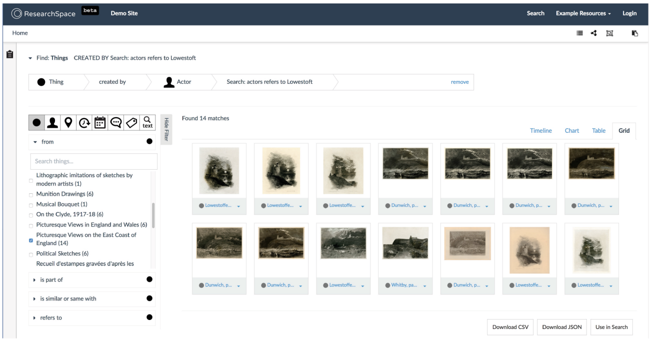

Fig. 3. A two part search that identifies first actors that refer to Lowestoft, and then the actual works created

```
@prefix rso:   <http://www.researchspace.org/ontology/>.
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#>.
rso:Actor_refers_to_Place rso:hasDomain rso:Actor;
rso:hasRange rso:Place;  rdfs:label "refers_to" .
```

Listing 1.1. FR definition for Actor refers to Place

Figure 3 presents the query formulated by the researcher where Actor and Place are connected by a fundamental relationship (FR) *refers to* (see Listing 1.1). We exemplify in Listing 1.2 how that is specified in the system and the SPARQL construct to create *refers to* connections between actors and places. The current FCs and FRs are defined to cover the underlying data, but FCs and FRs can be adapted to reflect different onto-epistemological paths for particular datasets and according to different theoretical approaches. For example, archaeological data might be used to answer anthropological, economic and social questions and carefully designed FCs and FRs could encourage particular perspectives and approaches useful for different types of research. Therefore, this component is not simply a solution to the accessibility problem, but a component that incorporates expert rules for specialist data investigation.

```
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX rso: <http://www.researchspace.org/ontology/>
CONSTRUCT {?actor rso:Actor_refers_to_Place ?place .
               ?place rso:Place_is_referred_to_by_Actor ?actor}
WHERE {
        ?actor a rso:Actor .
        ?place a rso:Place .
        ?actor rso:Actor_is_creator_of_Thing
           / rso:Thing_refers_to_Place ?place}
```

Listing 1.2. SPARQL CONSTRUCT for inferring the FR Actor refers to Place

Structured Search with Knowledge Patterns In this case, search is about interrogating the structured data by using knowledge patterns i.e. SPARQL graph patterns rather than abstractions. However, it employs the same UI as graph driven searches using domain and range assignments avoiding the pre-processing of inferences. It allows researchers to apply specific knowledge patterns, useful in specialist projects.

Text-Based Search Complementary to the previous type of search, the textual data attached to the fundamental categories in the graph is indexed using Solr[21]. This complements the other search types and makes looking for known items in the graph a straightforward process.

4 Example Projects Using ResearchSpace

In this section, we highlight three very different projects in terms of research goals: archaeology, art history and history of science. The functionality incorporated by each has been determined by the scope of the projects, but they all share the same thinking in terms of digital scholarship. Explicitly, this refers to the sharing and reuse of data, the transferability of knowledge across contexts and over time, as well as its continuous enrichment, and to scholarly publications where data has the function to strengthen the argument. Early versions of ResearchSpace have already been deployed and used as the prototyping platforms for an archaeology and an art history project, while the third project has a live version of a customised instance of ResearchSpace. In each project the knowledge graphs are being grown to address different questions and accommodate different methods of research.

4.1 Archaeology: Geometric Reconstruction and Novel Semantic Reunification of Cultural Heritage Objects

This interdisciplinary project[22] aims to support reassociation of object fragments with shared characteristics (e.g. same school of production, style, age), the unification of object parts separated across collections, and, if possible, reassembly of fragments helped by algorithmic modeling. It employs a desktop application for 3D image analysis and annotations of 3D objects' representations and creates a knowledge graph in an integrated ResearchSpace instance pulling together data enriched by human agents and algorithmic analysis with explicit provenance.

The prototype application uses an existing body of information extracted from museums' inventories as a starting point for the knowledge graph. This is enriched with new information from 3D analysis and with assertions from the users. Using 3D scans of fragment edges, the algorithms developed within the project determine possible relations between fragments. The likely matches are assessed within the context of their historical data, and passed to humans for a

[21] https://demo.researchspace.org/resource/Help:SolrFullTextSearchSyntax.
[22] http://gravitate-project.eu.

final determination. The aim is to reduce the amount of time required for these operations overall, given a large number of dispersed fragments.

The organic expansion of the graph through continuous enrichment either from human or algorithmic agents, including structured argument, accompanied by knowledge provenance is paramount. It helps users understand the sources of the archaeological data and establish trust in the system. The prototype system has received positive and constructive feedback during a workshop session at the Computer Applications and Quantitative Methods in Archaeology (CAA) conference in March 2018.

4.2 Art History: Late Hokusai (British Museum/SOAS)

The focus of this project[23] was to bring together existing scholarly information about the work, life, and historical context of the Japanese artist Katsushika Hokusai into a searchable and explorable resource. Trained curators have undertaken the technical task of mapping to RDF using CIDOC CRM, custom scripts, and tools such as [12]. Hokusai and his works, namely his paintings, single-sheet impressions, illustrated books, drawings and letters from the British Museum (London) together with data from the Freer-Sackler Gallery (Washington DC), Metropolitan Museum of Art (New York) and the Art Research Center, Ritsumeikan University, have been semantically linked and enriched through transcriptions, translations and annotations. This enrichment applies to the primary material, and to secondary sources that document Hokusai's late period of activity in Japan. The Hokusai instance of ResearchSpace aims to be a multilingual research and knowledge platform, providing an alternative to current reference systems that restrict research input due to limited pre-defined categories. It provides new ways of investigating the archetypal representative of the ukiyo-e ('floating world') school with all its richness and complexity encoded in a specialist network of knowledge. Feedback on the development of the resource was received at workshops in Washington DC and New York (2018) with researchers expressing particular interest in the use of argumentation which is fundamental to various aspects of the project in particular connoisseurship.

4.3 History of Science: CorpusTracer (Max Planck)

This project uses ResearchSpace to investigates the knowledge networks and history surrounding one book: the Tractatus de sphaera of Johannes de Sacrobosco [10]. It is an example of a highly specialised digital monograph, but one which can be extended and integrated with other related digital resources and 'grown' into a network of interrelated knowledge. By applying methods from network analysis, it investigates how specific commentaries on the text circulated, which actors were responsible for them, and what factors supported

[23] http://gtr.rcuk.ac.uk/projects?ref=AH%2FN00440X%2F1.

or hindered the spread of specific kinds of knowledge. The core of this investigation is achieved through the construction of the Corpus Tracer, a graph-database that uses primarily the IFLA FRBRoo[24] standard, a specialisation of the CIDOC CRM developed by the International Federation of Library Associations and Institutions (IFLA) and the CIDOC CRM Special Interest Group. The structured search with knowledge patterns has proven valuable for exploratory searches of the specialist knowledge graph.

5 Conclusions

This paper describes the rationale and some of the in use capability of ResearchSpace, a Semantic Web knowledge oriented system that is designed to work in, or help transform, knowledge environments into collaborative, argumentative, digital scholarly spaces through the contextualisation of data using onto-epistemological processes for semantic modeling. It supports interdisciplinary research and is additionally underpinned by material culture representing world history through the products of social relations.

This type of knowledge, created directly by academics and subject experts, is not represented by existing modes of data dissemination, which although useful, are based on data created for a different rationale and purpose. This difference in design and purpose is directly linked to the benefits and value of the data. Its dynamic enables the creation of Semantic Web applications for people who are interested in collaborative research and knowledge building, or want better contextual engagement by placing things within historical and theoretical settings, not provided by raw Linked Data.

References

1. Alani, H.: Unlocking the potential of public sector information with semantic web technology. In: Aberer, K., et al. (eds.) ASWC/ISWC 2007. LNCS, vol. 4825, pp. 708–721. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_51
2. Atkins, D.: Revolutionizing science and engineering through cyberinfrastructure: report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure (2003). http://hdl.handle.net/10150/106224
3. Bechhofer, S., et al.: Why linked data is not enough for scientists. Future Gener. Comp. Syst. **29**(2), 599–611 (2013). https://doi.org/10.1016/j.future.2011.08.004
4. Capadisli, S., Guy, A., Verborgh, R., Lange, C., Auer, S., Berners-Lee, T.: Decentralised authoring, annotations and notifications for a read-write web with dokieli. In: Cabot, J., De Virgilio, R., Torlone, R. (eds.) ICWE 2017. LNCS, vol. 10360, pp. 469–481. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60131-1_33
5. Dean, K.: Towards a Historical-Materialist Science of Thinking, Chap. 18, pp. 53–73. Routledge, London (2014)

[24] https://www.ifla.org/node/10171.

6. Doerr, M.: The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI Mag. **24**(3), 75–92 (2003). http://dl.acm.org/citation.cfm?id=958671.958678
7. Doerr, M., Kritsotaki, A., Boutsika, K.: Factual argumentation - a core model for assertions making. JOCCH **3**(3), 8:1–8:34 (2011). http://doi.acm.org/10.1145/1921614.1921615
8. Ehrlinger, L., Wöß, W.: Towards a definition of knowledge graphs. In: SEMANTiCS (Posters, Demos, SuCCESS) (2016)
9. Gradmann, S., Hennicke, S., Tschumpel, G.: Modelling the scholarly domain beyond infrastructure. In: DHd 2016, p. 143 (2016)
10. Kräutli, F., Valleriani, M.: CorpusTracer: a CIDOC database for tracing knowledge networks. Digit. Scholarsh. Humanit. (2017). https://doi.org/10.1093/llc/fqx047
11. Kuhn, T., Dumontier, M.: Genuine semantic publishing. Data Sci. **1**(1–2), 139–154 (2017). https://content.iospress.com/articles/data-science/ds010
12. Marketakis, Y., et al.: X3ML mapping framework for information integration in cultural heritage and beyond. Int. J. Digit. Libr. **18**(4), 301–319 (2017). https://doi.org/10.1007/s00799-016-0179-1
13. Meghini, C., Bartalesi, V., Metilli, D., Benedetti, F.: Introducing narratives in Europeana: preliminary steps. In: Kirikova, M., et al. (eds.) ADBIS 2017. CCIS, vol. 767, pp. 333–342. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67162-8_33
14. Oldman, D., Doerr, M., Gradmann, S.: Zen and the Art of Linked Data, Chap. 18, pp. 251–273. Wiley-Blackwell (2015). https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118680605.ch18
15. Oldman, D., Doerr, M., de Jong, G., Norton, B., Wikman, T.: Realizing Lessons of the last 20 years: a manifesto for data provisioning & aggregation services for the digital humanities (a position paper). D-Lib Mag. **20**(7/8) (2014). http://dblp.uni-trier.de/db/journals/dlib/dlib20.html#OldmanDJNW14
16. Paulheim, H.: Knowledge graph refinement: a survey of approaches and evaluation methods. Semant. Web **8**(3), 489–508 (2017). https://doi.org/10.3233/SW-160218
17. Rockwell, G.: As transparent as infrastructure: on the research of cyberinfrastructure in the humanities. In: Online Humanities Scholarship: The Shape of Things to Come, pp. 1–20 (2010)
18. Schneider, J., Groza, T., Passant, A.: A review of argumentation for the social semantic web. Semant. Web **4**, 159–218 (2013)
19. Stewart, C.A., Simms, S., Plale, B., Link, M., Hancock, D.Y., Fox, G.C.: What is cyberinfrastructure. In: Proceedings of the 38th Annual ACM SIGUCCS Fall Conference: Navigation and Discovery, SIGUCCS 2010, pp. 37–44. ACM, New York (2010). https://doi.org/10.1145/1878335.1878347
20. Veltman, K.H.: Towards a semantic web for culture. J. Digit. Inf. **4**(4) (2004). http://journals.tdl.org/jodi/article/view/113