SUMMARY

Semantic Web technologies allow cultural heritage institutions to publish interconnected, interoperable data, with explicit semantics. The source of the data published by museums is often the basic metadata recorded in systems aimed at collection management. As a consequence, users are deprived of the curated contextual information of regular exhibitions. To address this problem and provide better access to online collections, institutions employ various approaches to improve object descriptions. Among them is crowdsourcing, a quick and inexpensive source of large quantities of descriptions. However, it remains a challenge to ensure the quality of crowdsourced information, especially for knowledge-intensive tasks. In this thesis, we introduce nichesourcing, a method to solve knowledge-intensive tasks, by identifying and engaging small groups of experts. We present a five-step method, to enrich and contextualize object metadata using nichesourcing, thereby improving access to online cultural heritage collections.

ANALYSIS OF COLLECTION DATA

The first step of the method concerns the analysis of collection data. During this step, we assess the suitability of the chosen data model and the number of references to external datasets. In Chapter 2, we analyze the Linked Data of the Rijksmuseum Amsterdam, as a case study. The Rijksmuseum collection comprises over a million objects, of which only a fraction can be on display at a given time. To open up the remaining collection, the museum started to digitize objects and publish the resulting information online. The Linked Data of the museum consists of over 22 million statements, describing over 350,000 objects, of which more than 207,000 include a reference to an image. The data is used to support search, recommendation, collection integration and browsing.

The Rijksmuseum uses contextual concepts from structured vocabularies to describe objects. While the museum maintains its own vocabularies to preserve its own perspective, an increasing number of contextual concepts is related to external datasets. The collection data is structured using the Europeana Data Model. Not all aspects of the collection can be captured with the modeling constructs recommended by Europeana. Therefore, we discuss modeling challenges and proposed solutions for contextualizing cultural heritage data in the next chapter.

CONTEXTUALIZATION OF CULTURAL HERITAGE DATA

Ontologies make the semantics of data explicit, by providing a shared con-
ceptualization. When a cultural heritage institution wants to publish Linked Data, it is confronted with the choice of which ontology to use. This decision has implications for the source data that can be included, as well as the structure of the resulting Linked Data. As part of the five-step method, we focus in Chapter 3 on how ontologies can be used to structure and represent contextual information about objects in cultural heritage collections. We discuss modeling challenges that regard specialization, object- and event-centric approaches, temporality, representations, views and subject matter. For each challenge, we show modeling approaches of two ontologies often used in the cultural heritage domain: the Europeana Data Model and the CIDOC Conceptual Reference Model.

Based on the discussed modeling challenges, we formulate six requirements for cultural heritage ontologies: 1) the ability to specialize an ontology without decreasing its interoperability, 2) support for recording both attributes as well as events related to objects, 3) ability to capture changes over time, 4) ability to separate descriptions of objects and their representations, 5) support for capturing multiple sources describing the same object and 6) possibility to contextualize objects using subject matter. By considering these requirements, institutions can make a more informed choice when deciding on which ontology to use to contextualize data published online.

**Nichesourcing** The usefulness of cultural heritage data hinges on the quality and diversity of descriptions of collection objects. In many cases, existing descriptions are insufficient for retrieval and research tasks, resulting in the need for additional annotations. Eliciting such annotations is a challenge, since it often requires domain-specific knowledge. Where crowdsourcing can be successfully used to execute simple annotation tasks, identifying people with the required expertise might prove challenging for more complex and domain-specific tasks. Nichesourcing addresses this problem, by tapping into the expert knowledge available in niche communities.

In Chapter 4, we present Accurator, a methodology for conducting nichesourcing campaigns, by addressing communities, organizing events and tailoring a web-based annotation tool to a domain of choice. The contributions are the following: 1) a nichesourcing methodology, 2) an annotation tool for experts, 3) validation of the methodology in three case studies and 4) a dataset including the obtained annotations. The case studies concern birds on art, bible prints and fashion images. We compare the quality and quantity of obtained annotations, showing that the nichesourcing methodology in combination with the image annotation tool can be used to collect high-quality annotations in a variety of domains. A user evaluation indicates the tool is suited and usable for domain-specific annotation tasks.
**Summary**

**Diversification of Search Results**  In Chapter 5, we consider whether, and to what extent, additional semantics in the form of Linked Data can support explorative search. As a case study, we use the Linked Data of the Rijksmuseum, extended with various structured vocabularies. We apply an existing graph search algorithm to this data, which finds paths in the graph from the search term to target objects. Next, the algorithm clusters results with similar paths together. We use the number of resulting clusters and the path length as indicators of diversity. As sample queries, we collected the terms in the museum’s query log for the duration of one month.

The results show that for this application domain, the added semantics lead to 1) an increase in the number of results, and 2) an increase in the variety of search results. We hypothesize that the following two factors impact the usefulness of vocabularies for search: 1) the number of links between distinct concepts and the metadata objects and 2) the richness of the internal links between concepts in vocabularies. This fourth step of the method illustrates that additional semantics provided by structured vocabularies can help users to explore collections and reach more objects related to their interest.

**Integration of Collections**  Online cultural heritage collections often contain complementary objects, which makes integration of heterogeneous collections a worthwhile effort. In Chapter 6, we present the DigiBird system that provides access to four distinct nature-related collections and reinforces crowdsourcing initiatives. The system is designed to harmonize complementary collection objects, make crowd contributions instantaneously available and allow the monitoring of multiple crowdsourcing systems using one dashboard.

Harmonizing data from multiple systems that adhere to different standards proved to be a challenge. The data originates from dynamic systems, as a continuous stream of crowd contributions alters and extends the datasets. With the DigiBird system, institutions can decide to use data in an early stage of the crowdsourcing process. Additionally, undertaking crowdsourcing projects together allows the sharing of resources and provides insights into the time needed to collect results. By leveraging standardized data models and annotations from structured vocabularies, the DigiBird system illustrates the added value of enrichments and the benefits of Linked Data for collection integration.

**Conclusion**  In this thesis, we present a method to contextualize and enrich cultural heritage collections, to support explorative search and collection integration. Museums with similar collections as the Rijksmuseum will be able to use the method without requiring major changes. We expect that many steps of the method can be utilized in other domains as well. Disseminating high-quality information about
objects is embedded in the mission of cultural heritage institutions. Institutions that want to publish rich contextualized data online, have to assess the quality of external structured vocabularies and find efficient approaches to relate collection data to these new sources. Nichesourcing is one of these approaches, additionally providing ways to engage with the public. To keep contributors motivated, it is essential that they know their work matters. We show the direct impact of annotations on search functionality and collection integration, highlighting the potential of crowd contributions and Linked Data.