

## D-Lib Magazine

May/June 2014

Volume 20, Number 5/6

---

### Representing Cultural Collections in Digital Aggregation and Exchange Environments

Karen M. Wickett, University of Texas at Austin  
wickett@ischool.utexas.edu

Antoine Isaac, Europeana Foundation  
aisaac@few.vu.nl

Martin Doerr, FORTH-ICS, Crete  
martin@ics.forth.gr

Katrina Fenlon, University of Illinois at Urbana  
Champaign  
kfenlon2@illinois.edu

Carlo Meghini, Istituto di Scienza e Tecnologie  
dell'Informazione  
carlo.meghini@isti.cnr.it

Carole Palmer, University of Illinois at Urbana  
Champaign  
cpalmer@illinois.edu

doi:10.1045/may2014-wickett

---

#### Abstract

The representation of collections in digital library systems that aggregate or exchange cultural heritage data can serve a number of useful functions. In this article, we present specific roles that collections can play in digital aggregations, representational requirements that arise from those roles, and modeling strategies for meeting the requirements. The functional roles of collections and collection descriptions speak to the needs of individual users accessing or contributing content, system developers seeking to improve search experiences, and institutions providing data to federated aggregations. However, the current data models that support cultural heritage aggregations are not designed to fully accommodate and integrate collection-level data. Therefore we have developed a set of general requirements for the representation of collections in digital aggregation systems. In order to demonstrate how these requirements can be addressed in a current operational context, we present specific strategies for collection representation in systems that use the Europeana Data Model.

---

#### 1. Introduction

Collection structures and descriptions provide a variety of useful functions for users and managers of digital libraries, including technical capabilities for retrieval and evaluation of content, especially within large digital environments that aggregate many collections. Members of the [JMLS Digital Collections and Content \(DCC\) project](#), hosted by the Center for Informatics Research in Science and Scholarship at the University of Illinois at Urbana-Champaign, and developers of the [Europeana Data Model \(EDM\)](#) recently formed a collaborative study group to recommend an extension of EDM to explicitly accommodate representation of collections and collection/item relationships. The key findings of the collaboration are a set of roles that collections and collection descriptions can play in digital aggregation and exchange environments, the representation requirements that arise from those roles, and modeling strategies for meeting those requirements (Wickett, *et al.*, [2013](#)). Although the modeling recommendations are targeted at extending the EDM, the roles and requirements provide a general framework for collection-level representation and description in digital repositories, federated aggregations, and any systems that exchange cultural heritage data.

[Europeana](#) is a digital library aggregation system that provides access to digitized cultural heritage content from around Europe. While many of Europeana's data providers maintain collection-level entities or descriptions (e.g. [The European Library](#) and the [European Film Gateway](#)), Europeana itself does not currently use or preserve collection-level information. The primary goal of the collaboration between EDM and DCC was to examine the technical requirements for preserving, reconstructing, and building collection-level entities within the Europeana

context.

The Europeana Data Model (EDM) is the schema underlying Europeana's data ingestion, management, and publication. EDM aims to standardize representation of heterogeneous records while supporting:

- the description of digital resources and data ingestion processes separately from those for the description of original cultural objects,
  - the retention of complete item descriptions from data providers,
  - data enrichment by Europeana and third parties, leading to multiple records for the same object,
  - the description of complex objects, and
  - linking objects to other resources (concepts, places, persons...) related to them.
- 

## 2. Functional Roles of Collections and Collection Description in Aggregation Scenarios

Collections are an important aspect of institutional identity for the organizations that invest in their curation, digitization and public access. Collections are also a fundamental feature of information organization systems, providing technical capabilities for retrieval and evaluation of content within large aggregations. Perhaps most importantly, collection structures provide the organizational and intellectual context important to users for interpreting the relevance and significance of individual items for their purposes (Palmer, Zavalina & Fenlon, [2010](#)).

Collection-level entities and collection descriptions support a range of functions, including:

- representing data providers,
- providing context for items,
- managing and presenting search results,
- assessing relevance and accessibility,
- supporting contribution of collections by users.

Collections can only contribute in meaningful ways to the functionality of digital aggregations and exchange environments if information about those collections is made available within the system. In other words, the data models that support aggregation systems must be ready to fully accommodate collections and collection description.

Although collection description has received attention in the past from metadata researchers and model developers (e.g. Lagoze & Fielding, [1998](#); Heaney, [2000](#); Lee, [2000](#)), current data models and associated ontologies tend to be oriented totally around individual items and do not provide classes and properties that are sufficient to meet the representational requirements that arise from the potential roles for collections in aggregation scenarios. In particular, the current data model for Europeana, one of the largest and most influential cultural heritage digital aggregation systems, does not support collection description and representation.

---

## 3. Representational Requirements

The study group analyzed the functional roles that collections can play in aggregation scenarios in order to develop a set of representational requirements for data models to fully accommodate collections. Given the potential for the representation and descriptions of collections to improve the functionality of digital aggregations, it is essential for the underlying technical models to meet these requirements.

The following requirements for modeling collections correspond to the roles of collections listed above:

1. Models must treat collections as individual resources within the aggregation and allow for the representation of properties of the collection.
2. Models must be prepared to represent collection membership as a property that stands between resources. Item-level entities must be explicitly linked to collection-level entities.
3. It is necessary to have a set of properties designed to describe collections in ways that support users and managers. Properties essential to the use of collections in digital libraries and aggregations include:
  - a. Properties that record the institutions that have participated in the stewardship of resources; including

institutions collecting and/or holding physical resources, institutions that host digital versions of resources, and institutions that have created descriptions of resources.

- b. Properties that can be used to reflect the contextual information implied by collection membership, including topical or subject properties, properties related to the principles used to determine membership in the collection, and properties about the intended audience for a collection.

4. To the extent possible, property values in metadata should be identifiers of resources that the system can make actionable.

The requirements are intended to inform data model and schema development for digital aggregation and exchange systems, and are therefore very general. The next section discusses specific strategies for meeting these requirements in the case of Europeana or in any aggregation systems that use EDM.

---

## 4. Modeling Strategies for Collection Representation and Description

Following the overall goal to develop mechanisms for collection representation and description that can function to extend the Europeana Data Model (EDM), the strategies discussed below all adhere to the core EDM. Specifically, EDM extensively relies on the RDF modeling principles of using identifiable resources and statements for representing information about entities. This choice answers the final requirement above, but fully meeting that requirement will also rely on the provision of identifiers (especially, web identifiers) for any entity worthy of description, and the description of these entities as distinct resources.

The approach of the study group in determining the classes and properties needed to represent collections was twofold: (a) build on progress made on collection representation in the IMLS DCC project (Shreeves & Cole, [2003](#); Palmer, *et al.*, [2006](#)); (b) systematically align with the existing EDM classes and properties, or when such alignment is not possible, present new candidates as extension to the EDM. At the time of writing the technical report, EDM did not provide for expressing collections as resources with distinct properties and relationships. An EDM extension to this effect was desirable for the model to express data that meets the requirements presented above.

---

### 4.1 Defining the Class of Collections

EDM is designed to support integration of data from multiple sources, and the resources within the aggregation are represented as instances of classes as mentioned above. Therefore, in order to extend EDM to accommodate collections, the study group considered whether cultural collections could fit in the existing class hierarchy given by EDM, or whether it is necessary to introduce a new class into the model.

EDM prominently features three classes of resources:

- Provided Cultural Heritage Objects or CHOs (*edm:ProvidedCHO*) denote the original objects—either physical (e.g. a painting, a book, etc.) or born-digital (e.g. a 3D model), which are the focus of description and search in Europeana. The choice in granularity of description chosen for the ProvidedCHO belongs to the data provider, within the limits of relevance set by Europeana.
- Web Resources (*edm:WebResource*) represent digital representations of the provided CHOs, published on the web.
- Aggregations (*ore:Aggregation*) group the Provided CHO and the Web Resource(s) into one bundle, where information on the aggregation process is also recorded (e.g. the provider of the data).

EDM also defines contextual resources that can be used to provide more information related to the object (e.g. *edm:Agent*, *edm:Place*, *edm:Concept*, *edm:TimeSpan*).

In EDM, Aggregations are also used as context to create perspectives on CHOs ("proxies") that carry provider-specific data on these objects, thus allowing one to separate it from data on the same object from other providers (including Europeana). Therefore *ore:Aggregation* is primarily used in the model to serve as an organizing construct for repository managers and to aid in interoperability by providing assistance for harvesting or integration.

Representing collections as instances of the *Provided Cultural Heritage* (*edm:ProvidedCHO*) class adopts the standard object modeling methodology within Europeana since this class "comprises the Cultural Heritage objects that Europeana collects descriptions about" (Europeana Project, [2013](#)). Generally, the instances of this class are the main focus of the digitization and access efforts. Then, in the Europeana context of operation, the collection would

be embedded in an *ore:Aggregation*, which bundles the collection with its digital representation(s), including its homepage, for example.

Since *edm:ProvidedCHO* is a functional class that does not constrain the exact nature of resources, a collection simply typed as a Provided CHO would be difficult to distinguish from its item-level members, also typed as Provided CHOs. A candidate to reflect the intended semantics for a class of collection is *dcmitype:Collection*, an element of the DCMI Type Vocabulary provided by the [Dublin Core Metadata Terms](#), defined as "an aggregation of resources."

However, it is problematic to directly re-use *dcmitype:Collection* in contexts that use *ore:Aggregation* for technical purposes, because *ore:Aggregation* is defined as a subclass of *dcmitype:Collection*. This means that in systems that use subclass reasoning, a query for resources of type *dcmitype:Collection* will return all resources of type *ore:Aggregation*, which is problematic given the use of *ore:Aggregation* to manage varied representations of objects in EDM. In addition, the very general definition of *dcmitype:Collection* includes any given set of resources, a scope that is considerably broader than the one of intentionally created or curated collections. Therefore, the study group has proposed defining the class *edm:Collection* as a subclass of *dcmitype:Collection*, with the definition "a group of objects gathered together for some intellectual, artistic, or curatorial purpose."

---

## 4.2 The Collection Membership Relationship

In order for collections to play their expected role in digital library aggregation and exchange environments, collection membership must be represented as a property that stands between resources. This property can then be used to explicitly link item-level entities to the collection-level entities of which they are members. This kind of linking will not be possible in aggregation or exchange scenarios where items are not given individual representation, but where items are available, the explicit representation of the membership relationship is an essential element for supporting the roles of collections listed above.

The [DCMI Metadata Terms](#) defines *dcterms:hasPart* as "A related resource that is included either physically or logically in the described resource", and *dcterms:isPartOf* as "a related resource in which the described resource is physically or logically included." Since an item is logically included in a collection that it has been gathered into, these terms are appropriate for representing collection membership. However, these parthood relations may be too general for the representation of collection membership in digital library aggregation and exchange environments. There are many kinds of parthood relations that may be represented with *dcterms:hasPart*. For example, pages are parts of books, and volumes are parts of series, and these seem like semantically distinct relationships from collection membership. It is perhaps most accurate to characterize collection membership as a particular kind of parthood.

A strategy that maintains a connection to the commonly used Dublin Core property while indicating specialized semantics for collection membership is to define a new property, *edm:isGatheredInto* specifically for collection membership as a sub-property of *dcterms:isPartOf*. The sub-property relationship means that every instance of *edm:isGatheredInto* implies a corresponding instance of *dcterms:isPartOf*. This connection from the specialized collection membership relation to the more general parthood relation will support interoperability between different applications.

---

## 4.3 Collection-level Description

The usefulness of collections in large-scale digital aggregations depends on collection-level description. Collections must be described according to a collection-level schema for users to find and identify them as information objects, or for the managers of aggregations to use them to represent the contributions of data providers. Whenever a resource is accessed by a user, the contextual information should be readily available and some elements of the context may be directly presented to the user, depending on the specific access function. Contextual properties include topical or subject properties, properties related to the purposes a collection was created to serve, and properties about the intended audience for a collection.

The [DCC Collection Description Metadata Schema](#) is a data structure aligned with the [Dublin Core Collections Application Profile](#) that is designed for representation of collections as well as items, and the relationships between them. The study group analyzed how the existing schema fits the user requirements, and considered the connections between these collection-level properties and the roles and requirements. The full property analysis as presented in the technical report is intended to provide a starting point for the development of an EDM application profile for describing collections. This work builds on earlier developments in collection description from the digital library and

metadata fields (Heaney, 2000; Powell, Heaney, & Dempsey, 2000) and integrates recent perspectives and experience with semantic web and linked data approaches (Heath & Bizer, 2011) to produce recommendations for collection description that support users and administrators of current digital aggregation and exchange environments.

The alignment with EDM was realized by (i) mapping the DCC schema fields onto the available properties used by EDM, introducing extensions where necessary, and (ii) specifying the classes of resource the properties should be attached to. Following the recommendations from the study group discussed in Section 4.1, collection representation in EDM would result in two entities: one instance of both *edm:ProvidedCHO* and *edm:Collection* that represents the collection as an intellectual creation, and instance of *ore:Aggregation* that bundles the collection together with its digital representations (see Figure 1).

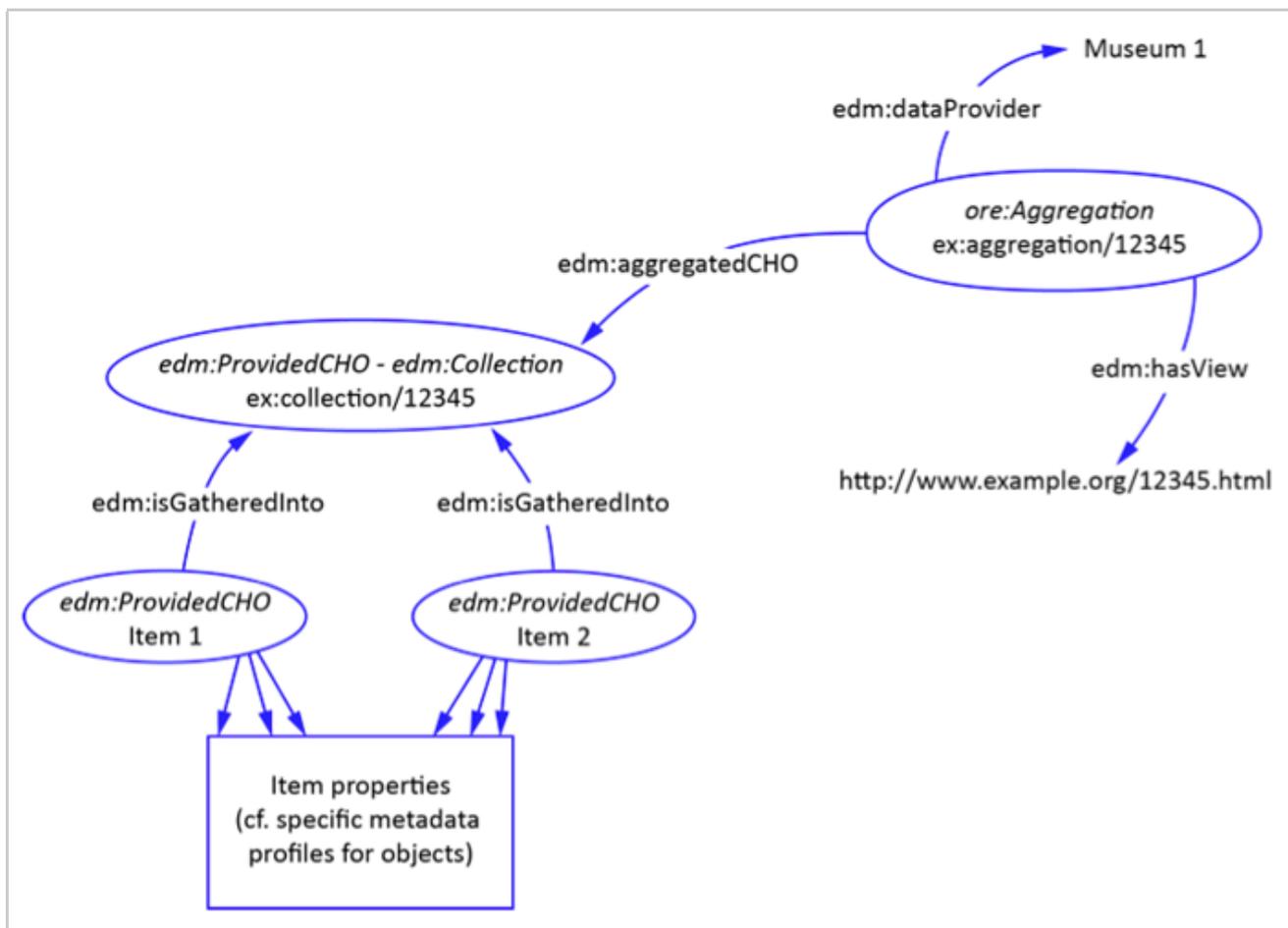


Figure 1: Core entities and properties for collection representation

The study group organized the collection-level properties into six categories:

1. **Collection identity properties** assign and manage properties of and relationships to individual collections. These properties meet the requirement of treating collections as individual objects, and provide for collection description generally. In the EDM context, these properties include *dc:title*, *dcterms:alternative* (for an alternative title of a collection), and any *dc:identifier* statements for a collection, and are attached to the instance of *edm:ProvidedCHO* that represents the collection itself.
2. **Access properties** support access to and presentation of digital representations of collections via the web. In addition to the functional role these properties play in access, they also represent the details of hosting digital representations of collections. In the EDM context, these properties include *edm:isShownAt*, and *edm:rights*, and are attached to the instance of *ore:Aggregation* that specifies the digital collection context of a given collection.
3. **Aggregator context properties** record information essential to the operation of data creation and aggregation systems. These properties meet the requirement to record information about the stewardship of collections, including institutions creating and hosting digital representations of collections. In the EDM context, these

- properties include *edm:provider* and *edm:dataProvider*, and are attached to an instance of *ore:Aggregation*.
4. **Collector context properties** reflect aspects of the creation of a collection via the gathering together of individual items, representing the intent of a curator or scholar with respect to the collection and facts about the collection process. These properties meet the requirement to record information that reflects the context implied by collection membership. In the EDM context, these properties include *dc:creator* (applied to the creator of the collection), *dcterms:extent*, *dcterms:accrualPolicy*, *dcterms:audience*, and *dc:description*. These properties are attached to an instance of *edm:ProvidedCHO* that represents a collection.
  5. **Secondary collector context properties** describe relationships between collections and reflect the embedding or inclusion of one collection-level entity into another collection-level entity. These properties include *dc:relation*, *dcterms:isPartOf* and *dcterms:hasPart*, and are attached to an instance of *edm:ProvidedCHO*.
  6. **Item-related properties** indicate attributes of the particular items that have been gathered into a collection. These properties give a more complete view of a collection, particularly in scenarios where item-level descriptions are not available for direct access in a repository. In the EDM context, these properties include *edm:itemCreator*, *edm:itemGenre*, *clid:dateItemsCreated*, *dcterms:spatial*, and *dcterms:temporal*, and are attached to an instance of *edm:ProvidedCHO* that represents a collection. When the information represented by these properties is represented at the item level, it may be possible to derive the collection-level properties using inference rules based on relationships between collection and item metadata (Wickett, Renear & Urban, 2010).
- 

## 5. Conclusion

The representation of collections in digital cultural heritage aggregations and exchange environments has the potential to serve a range of intellectual, administrative, and functional roles. In order to meet this potential, the data models that support aggregation systems must be ready to fully accommodate collections and collection descriptions.

We have addressed the modeling need at two levels. The representational requirements derived from the roles of collections speak to the general needs for representing collections usefully, while the modeling strategies are intended to inform practice in implementation scenarios that use the Europeana Data Model. Therefore, while these recommendations can inform collection representation scenarios generally, they may be particularly useful for aggregation systems at regional, national, or international levels that have data models based on the EDM.

Collection representation is currently an active area of research in digital libraries. The full technical report produced by the study group (Wickett, *et al.*, 2013) gives greater detail on the analysis of collection roles and representation and discusses some further areas for research and development, including the intellectual nature collections and the criteria that bind collection members into coherent wholes, and rules for the propagation of information between collections and items.

---

## Acknowledgements

The planning and development of the collaborative study group was funded by the IMLS Digital Collections and Content project (DCC), Principal Investigator, Carole L. Palmer, Center for Informatics Research in Science and Scholarship (CIRSS). The study group benefited from the participation of Allen H. Renear, David Dubin, Jacob Jett and Megan Senseney.

---

## References

- [1] Europeana Project (2013). [Definition of the Europeana Data Model](#). Version 5.2.4.
- [2] Heaney, M. (2000). [An analytic model of collections and their catalogues](#). *UK Office for Library and Information Science*.
- [3] Heath, T. and Bizer, C. (2011). [Linked Data: Evolving the Web into a Global Data Space](#). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
- [4] Lagoze, C. and Fielding, D. (1998). Defining collections in distributed digital libraries. *D-Lib Magazine*, 4(11). <http://doi.org/10.1045/november98-lagoze>

- [5] Lee, H. (2000). [What is a collection?](#) *Journal of the American Society for Information Science*, 51(12), 1106-1113.
- [6] Palmer, C. L., Knutson, E., Twidale, M., & Zavalina, O. (2006). Collection definition in federated digital resource development. In *Proceedings of the 69th ASIS&T Annual Meeting (Austin, TX)*. <http://doi.org/10.1002/meet.14504301161>
- [7] Palmer, C. L., Zavalina, O., & Fenlon, K. (2010). Beyond size and search: Building contextual mass in aggregations for scholarly use. *Proceedings of the American Society for Information Science & Technology*, 47. <http://hdl.handle.net/2142/18655>
- [8] Powell, A., Heaney, M., and Dempsey, L. (2000). RSLP Collection Description. *D-Lib Magazine*, 6(9), 1082-9873. <http://doi.org/10.1045/september2000-powell>
- [9] Shreeves, S. L. and Cole, T. W. (2003). [Developing a collection registry for IMLS NLG digital collections](#). In *Proceedings of the 2003 international conference on Dublin Core and metadata applications: supporting communities of discourse and practice—metadata research & applications* (DCMI '03).
- [10] Wickett, K. M., Isaac, A., Fenlon, K., Doerr, M., Meghini, C., Palmer, C. L., and Jett, J. (2013). "Modeling Cultural Collections for Digital Aggregation and Exchange Environments." CIRSS Technical Report 201310-1, University of Illinois at Urbana-Champaign. <http://hdl.handle.net/2142/45860>
- [11] Wickett, K. M., Renear, A. H., and Urban, R. J. (2010). Rule categories for collection/item metadata relationships. In *Proceedings of the 73rd ASIS&T Annual Meeting (Pittsburgh, PA)*. <http://doi.org/10.1002/meet.14504701218>
- 

## About the Authors



**Karen M. Wickett** is an Assistant Professor in the School of Information at the University of Texas at Austin. Her research is on the conceptual and logical foundations of information organization systems and artifacts. She is interested in the analysis of common concepts in information systems, such as documents, datasets, digital objects, metadata records, and collections.

---



**Antoine Isaac** works for Europeana.eu, a vast network for providing access to cultural heritage material in Europe's libraries, archives and museum. There, he is responsible for R&D collaborations, especially involved in data modeling and exchange. He is also a guest at the Free University of Amsterdam. His background is in Semantic Web research and applications. He has been involved in a number of W3C-related activities especially on Semantic Web Deployment (mostly for the vocabulary SKOS, for which he still acts as 'community contact') and Library Linked Data.

---



**Martin Doerr** s Research Director at FORTH-ICS in Heraklion, Crete. He has been leading or participating in a series of national and international projects for knowledge management, terminology management, cultural information systems and information integration systems. He is leading the working group of ICOM/CIDOC (International Committee for Documentation of the International Council of Museums) which has developed ISO21127:2006, together with the respective ISO committees, a standard core ontology for the semantic interoperability of cultural heritage information and beyond. His research interests are ontology engineering, semantic interoperability and information integration.

---



**Katrina Fenlon** Katrina Fenlon is a doctoral student in the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. As a research assistant in the Center for Informatics Research in Science and Scholarship, she investigates organizational, descriptive, and contextual problems in digital libraries of humanities and cultural heritage collections.

---



**Carlo Meghini** is prime researcher at CNR-ISTI in Pisa since 1984. He graduated in Computer Science at the University of Pisa in 1979 with a research thesis on distributed databases. He is involved in European projects since 1988, in the areas of Multimedia Information Retrieval, Digital Libraries and Digital Preservation. From 2007 he has contributed to the making of Europeana, the European digital library, taking care of the scientific aspects, notably the specification of the Europeana Data Model.

---



**Carole Palmer** is a Professor in the Graduate School of Library and Information Science and Director of the Center for Informatics Research in Science and Scholarship (CIRSS) at the University of Illinois at Urbana-Champaign. Her research investigates how to advance large-scale digital research collections, especially for interdisciplinary inquiry. At CIRSS she leads data curation initiatives focusing on the reuse value of data, scholarly use of digital collections, and data curation education and workforce development.

---

Copyright © 2014 Karen M. Wickett, Antoine Isaac, Martin Doerr, Katrina Fenlon, Carlo Meghini and Carole Palmer