Managing Research Data Lifecycles through Context

Presenters:

Grace Agnew, Associate University Librarian for Digital Library Systems

Ryan Womack, RUcore Research Data Manager

The Rutgers University Libraries have taken a unique and extensible approach to supporting research data lifecycle management at the Rutgers University Libraries. This approach involves leveraging the existing RUcore institutional repository, while extending it to support research data, as well as redesigning positions and developing a team to provide the expertise and support needed by busy researchers.

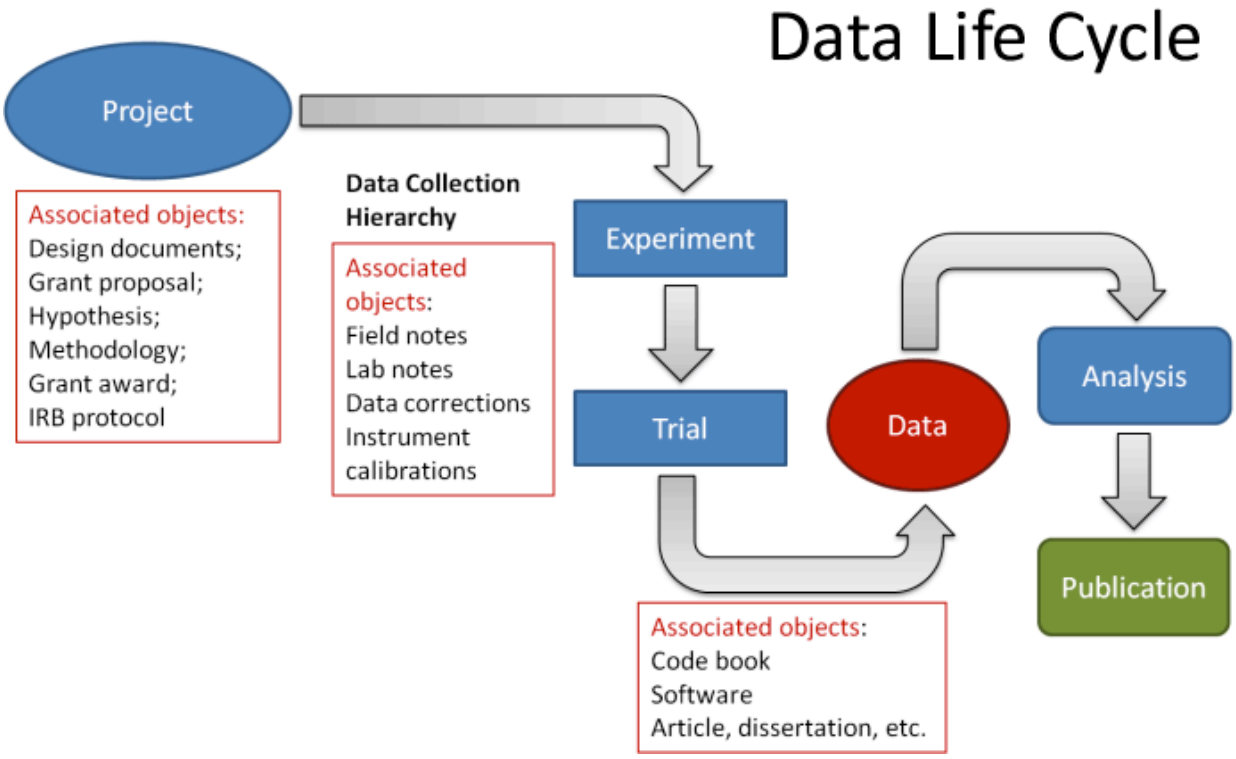
The Technical Infrastructure

RUcore, the Rutgers Community Repository , is a suite of tools and services built upon the open source Fedora Commons repository software. The Rutgers University Libraries developed applications to support digital resources of all types, which are available as open source, to support the discovery, access and reuse of digital resources. These applications include the faculty publications submission service, the RUetd electronic theses and dissertations submission service and RUresearch, the Rutgers research data portal. The heart of the RUcore service suite is the Workflow Management System, available for download as the OpenWMS, which provides a unique descriptive utility (metadata) as well as digital resource handling and management for linking resources to metadata and integrating them into the repository collection.

Research resources are defined as any resource created during the research process, which may be data sets, video, audio, images, lab notes, etc. Data produced during the research process is dense and rich. Many resources are created, and should be brought together, during the research process to support interepretation of the results of the research, reuse of research products, replication and validation of research, as well as analysis and interpretation of research. And all this needs to occur in a multi-disciplinary context, so that economists can reuse environmental data, political scientists can benefit from educators' reasoning analysis, etc.

A typical grant-funded research project might include resources from the grant process, such as the proposal, IRB protocol, project plan, annual reports, as well as resources from the research itself, which might be data collected from multiple experiments and multiple trials within each experiment. Many ancillary resources, such as images, data correction documentation, field or lab notes, etc. may also be collected and each of these may be created at distinct points in time, for distinct reasons, by different researchers. Equally important is to capture the products that interpret and validate the data, as well as showing impact for the data—the publications, such as dissertations and articles, or instances of reuse of the data by the same or other researchers. This is rich information that is generally stored in many places, from document/project management platforms such as Zotero or Sakai to institutional

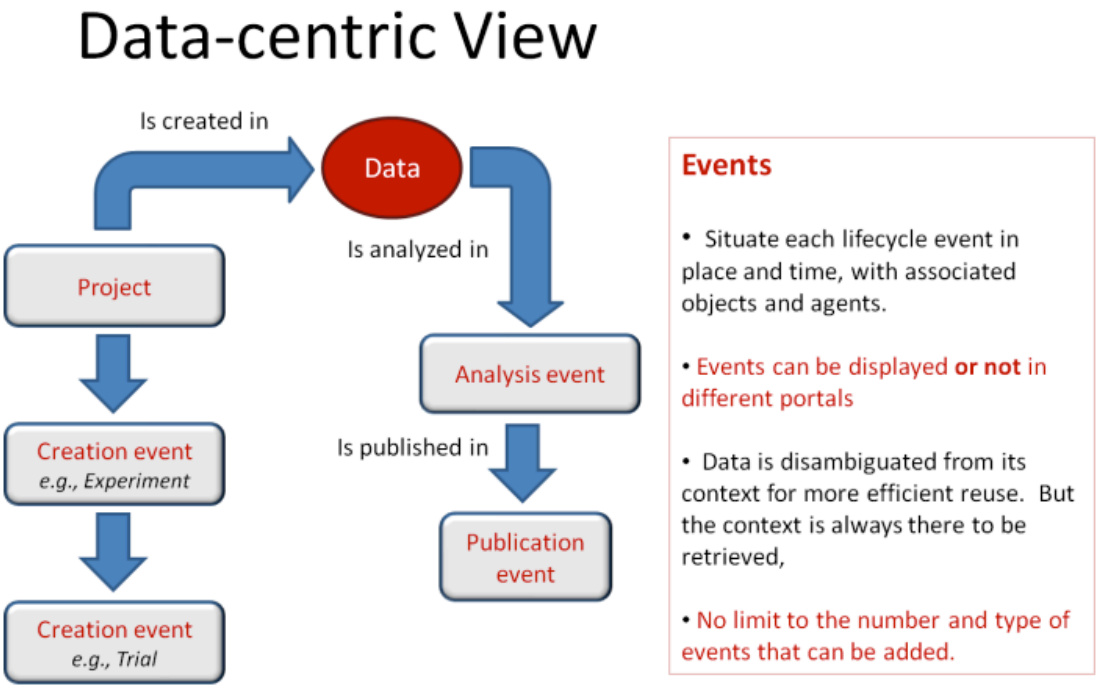
repositories, to the hard drives of the many different participants in the project. In the process of storing this information, which occurs simultaneously with the research itself, important context (who, what, when, where, and why) and relationship (this preceded that, this was revised because of that) is often lost. Anyone who has attempted to build forward upon or recreate previous research has discovered that it is not enough to simply bundle everything together. Meaningful context that can describe the provenance and relate resources to the research ecosystem is imperative.



The Rutgers University Libraries establish context through a unique METS (Metadata Encoding and Transmission Standard) application profile. METS is a library standard that enables the capture of metadata about the content of a resource (descriptive), it's provenance (source); its technical characteristics (technical) and the rights associated with its use (rights). The RUcore application profile includes events that capture the **what** event occurred in the lifecycle of the resource, **who** was involved in the event, **what related resources** are associated with the event and **when/where** did the event occur. Other resources can be related to the event both through a link to the other resource but also through the descriptive information included in the event, which explains the context of what occurred.

Rights Event Example *Equine Science Center Videos*	Data Life Cycle Event Example
Rights Event **Type** Permission of License **Date Time** 2009-12-09 **Associated Entity** **Role** Contributor **Name** Elyse Conway **Associated Object** **Type** Publicity release **Name** Model release **Detail** rutgers-lib:27494	**Data Life Cycle Event** **Type** Related publication **Label** Article references the data set, Major League Baseball and Performance Data, 1986 **Name** Pazzani, Michael J. and Bay, Stephen D. (1999) The Independent Sign Bias: Gaining Insight from Multiple Linear Regression in Proceedings of the Twenty First Annual Conference of the Cognitive Science Society. **Identifier** http://hdl.rutgers.edu/1782.1 /rucore30016700001.Manuscript.000056844

Because events are defined and documented in a standardized manner, using a controlled vocabulary for event type, events can be selected for display in different portals for different purposes. Events can also be hidden from public display when they are used to manage the resource rather than to support discovery and reuse. Examples include a rights event for an IRB protocol which is used to associate the protocol document closely with an object so that it can be reviewed as needed to ensure protection of human or animal subjects and a license event which documents the rights that the rights holder has provided to the RUcore repository. The ability to categorize and treat events differently enables the creation of a large number of events, without creating unnecessary "noise" for audiences with no interest or need to know specific events. This enables a data centric approach to managing data, where the data and its discovery and use are pre-eminent.



Management of Research Data

RUcore offers a very sophisticated platform for description, discovery and access to data. Custom portals can be created for different audiences, in addition to the central portal, RUresearch, which is designed to support multidisciplinary research data use. Working with faculty to develop data plans that utilize the tools of RUcore in an efficient manner requires a team of public and digital services staff who are engaged in working with faculty and their data. The Rutgers University Libraries are evolving a team, which is led at the administrative level by the Associate University Librarian for Digital Library Systems, who oversees the development of the RUcore platform and the technical staffing for the repository and its portals, including RUresearch. The team consists of members with expertise in repository programming and design, digital data curation, metadata, project management, intellectual property management and liaison librarians working directly with faculty. While each member of the team brings different expertise to working with data, it is critical that everyone share a core understanding of the nature of research data, how it is created, how users will work with data, and how it can be organized, managed and sustained in the RUcore repository. This requires a shared and strong understanding of the data model for research data projects so that all the entities created as part of the research process are captured, described and made accessible to researchers worldwide. Describing this data requires the ability to discover, design and implement vocabularies specific to the research community for the data, to develop robust and effective formats for storing and sharing the data, and to enable sharing the data with the wider research community through supporting or mapping the metadata or digital data formats in use by the community.

The Rutgers University Libraries are taking a two-pronged approach to creating and managing a team from which flexible project-specific teams can be assembled to work directly with faculty on their research data. An eleven session course, taken over seven months, will teach in-depth principles and practices as well as grounding everyone in the tools and technologies available via RUcore. In addition, members of the team will meet monthly to share expertise and progress for ongoing projects, to develop specifications for the ongoing development of RUresearch and to provide continuing education and professional development beyond the initial course. In addition, the team will continually evaluate its own organizational composition and effectiveness, identifying workflow bottlenecks, expertise holes, single points of failure, etc. This work will be used to recruit additional team participants and to design position vacancies to support a very strategic but still emerging area of library service.