



A Data Staging Repository for Digital Research Data

DataStaR is intended to facilitate collaboration among researchers, as well as the documentation and transmission of research data sets from a variety of disciplines to domain-specific repositories and/or institutional repositories. The data staging model leverages the ability of a researcher's local institution to provide accessible support and services related to research data, early in the research process, and by promoting the deposition of data in domain-specific repositories, makes data available to the larger research community.

Project team - Cornell University Library:

Gail Steinhart, Brian Caruso, Kathy Chiang, Jon Corson-Rikert, Dianne Dietrich, Brian Lowe, Janet McCue

Collaborator:

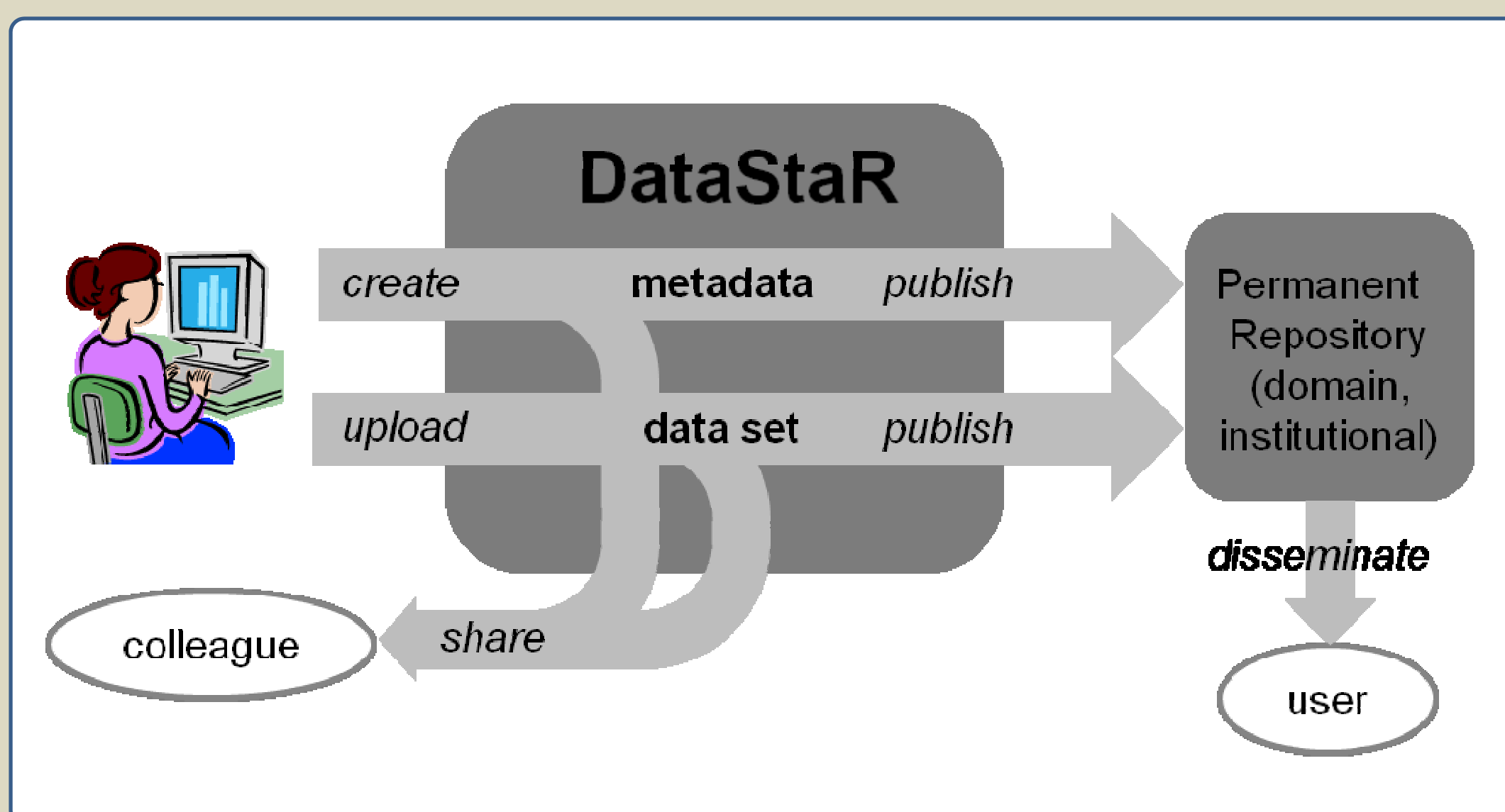
Ann Green, Digital Life Cycle Research and Consulting

Contact - Gail Steinhart GSS1@cornell.edu

Project web site: <http://datastar.mannlib.cornell.edu/>

DataStaR is a platform:
(in development)

- DataStaR is an online repository where researchers can share data with selected colleagues.
- DataStaR is a database of information about participating researchers and research groups.
- DataStaR allows users to create metadata in a variety of formats, reusing information from the researcher database and from their own metadata records.



DataStaR is a set of services:

DataStaR staff can help researchers select a repository with which to publish their data, determine the data and metadata requirements for repositories, and help researchers meet those requirements.

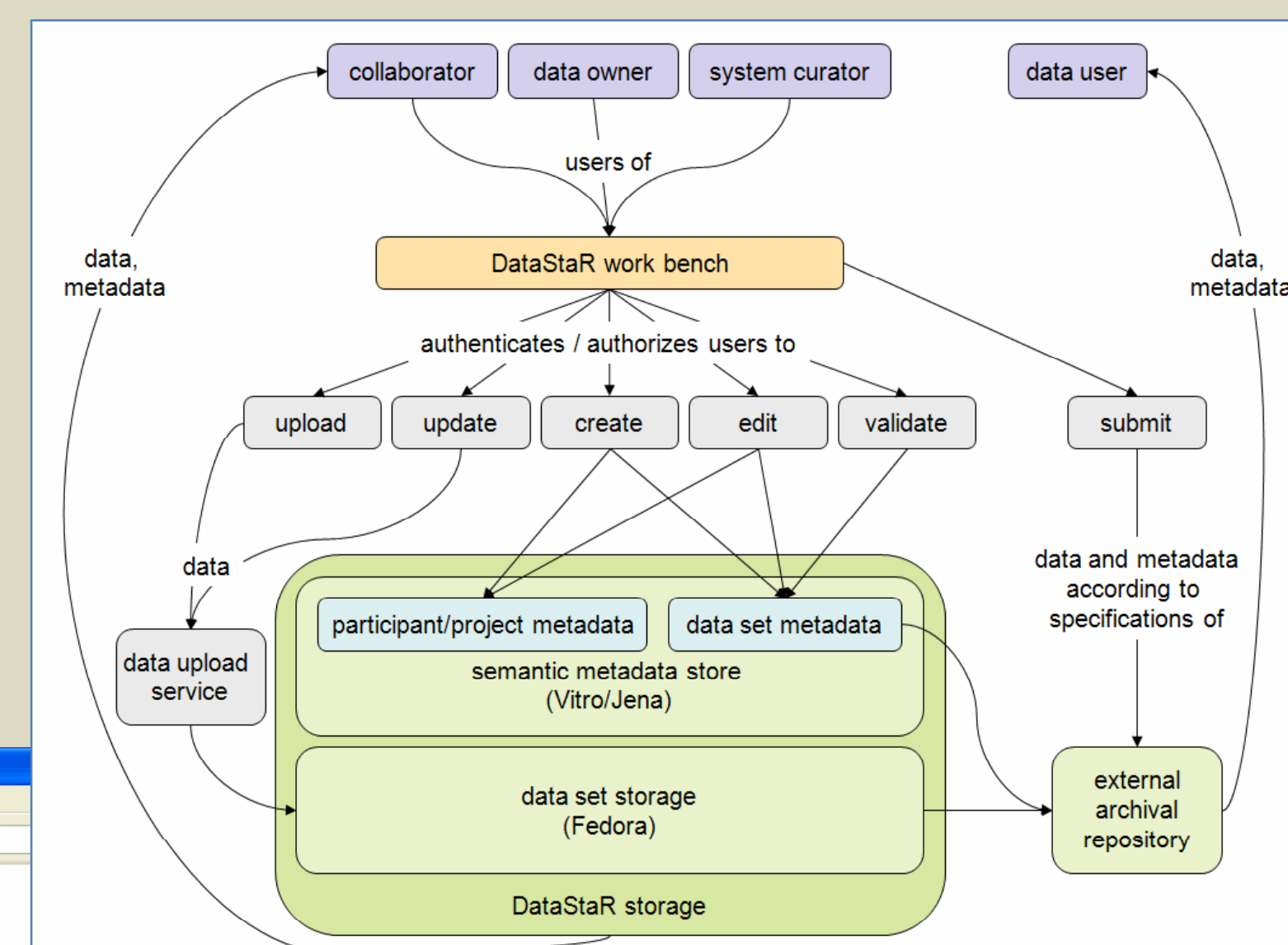
A semantic web approach to managing metadata:

- extend **vitro**, a Java web application that provides a customizable front end for searching and browsing a semantic graph of data, along with an interface for editing ontologies and instances (<http://vitro.mannlib.cornell.edu/>)
- “lift” existing metadata schemas into OWL ontologies
- treat metadata as a collection of discrete statements
- store and harness information about users and research groups for metadata and for public display
- reuse and recombine statements to create new metadata without repetitive entry of information

#	Property	domain
1	acceptsMetadataStandard	Data repository or center
2	affiliation	Person
3	associatedWithResearchGroup	Data set
4	assocWithResearchGroup	Document
5	contributor	
6	contributorTo	(Person or Organization)
7	coPrincipalInvestigatorOf	Person
8	creator	
9	creatorOf	Person
10	dataSetWillBeSubmittedTo	Data repository or center
11	formatOfFile	File format
12	fundedBy	Research group
13	fundsResearchGroup	Organization
14	hasAffiliate	Organization
15	hasContributor	Document
16	hasCoPrincipalInvestigator	Research group
17	hasCreator	Creatable Thing
18	hasDataSet	Research group
19	hasFileFormat	File
20	hasOfficialAffiliate	Organization
21	hasOfficialAffiliation	Person
22	hasParticipant(s)	Research group
23	hasPrincipalInvestigator	Research group
24	hasPublication	Research group
25	hasRelatedDataSet	Topic
26	hostsDataSet	Data repository or center
27	managedByOrganization	Data repository or center

Partial list of core DataStaR ontology properties

Public display of information about a participating research group



Architecture: Fedora for repository storage; DataStaR/vitro for metadata

Participating research groups:

- Cayuga Lake Watershed Network
- Cornell Biological Field Station
- Cornell Plantations Natural Areas
- Cornell Language Acquisition Laboratory
- Upper Susquehanna Agricultural Ecology Program

Preliminary target repositories and metadata standards:

- eCommons@Cornell (DSpace/Dublin Core)
- Cornell University Geospatial Information Repository (FGDC Content Standard for Digital Geospatial Metadata)
- Knowledge Network for Biocomplexity (Ecological Metadata Language)