Member Node Description: FEMC (Forest Ecosystem Monitoring Cooperative)

Version 1.0 06/28/2017 Mike Finnegan, Jim Duncan

General

Name of resource:	Forest Ecosystem Monitoring Cooperative (FEMC)
URL(s):	httsp://www.uvm.edu/femc/
Institutional affiliation(s):	University of Vermont
Primary geographic location:	Northeastern United States
Project Director & contact info:	Jim Duncan, James.Duncan@uvm.edu, (802) 656-2975
Technical Contact & contact info:	Jim Duncan, James.Duncan@uvm.edu, (802) 656-2975
Age of resource:	Since 1990
Funding support:	US Forest Service Northeastern Area State and Private Forestry, Vermont Agency of Natural Resources, University of Vermont
Proposed Unique Identifier:	urn:node:FEMC

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):

The FEMC focuses on collecting and maintaining one of the longest and most expansive records of forest health, wildlife, soil, air, and water quality data in the region, which encompasses the northeastern US. Datasets range from a single collection point with just a few records to sampling sites located over several acres to long-term datasets spanning decades with by-the-minute granularity. Hosting data from federal, state, non-profit, professional, and academic institutions from around the northeast serves to provide the information necessary to conserve and manage forested landscapes, identify and monitor threats to forest health and function, and facilitate collaboration.

Types of data (complex objects, text, image, video, audio, other):

We store data primarily as simple data types in tables within our relational database (text, int, datetime, etc.). We also store data in native file formats of most any type, which includes documents (.xlsx, .pdf, .docx, etc.), images, compressed archives (including shapefiles), and raster files, among others.

Data and metadata availability (rights, licensing, restrictions):

Data and metadata are licensed for publishing in a layered approach. First, users can determine whether to publish metadata. Once metadata is published, they can select an option for publishing data. Data can be

- withheld/embargoed, with a general policy of a 3-year renewable embargo term;
- made available upon request;
- made available for download under a Creative Commons license.

Option for embargo (yes/no, duration):

Yes, we have a policy allowing an embargo for 3 years (renewable) for research data, and a permanent embargo for sensitive data such as locations of threatened/endangered species.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):
Currently around 475 datasets, but likely to increase as we are actively conducting outreach to bring in additional contributors. Most datasets (~75%) have fewer than 3000 records, but around 10% have greater than 100,000 records, the maximum being just over 700,000. The vast majority of datasets are composed of just one table or file.

Please describe recent usage statistics, if known, including information on annual data product downloads, annual number of users, annual number of data products used in publications:
For an eight-month period (March – November 2016), we recorded 4595 unique client IDs, 571 of which visited dataset pages. For the same period, we tracked 341 downloads and also delivered custom data requests outside of the system.

User interactions

How does a user contribute data? (what can be deposited, how are data prepared, are specific software required, documentation/support available)
Most commonly, data is ingested into the system through a web interface, which processes the data supplied in a CSV file. Additionally, the data is inspected to make a "best guess" at a few metadata variables, prompting the user for confirmation or correction. Other files can be simply uploaded and remote datasets linked to. All datasets can have their metadata populated, regardless of where the data is stored or what format it is in. In addition, FEMC works with cooperators and contributors to ingest their data into the system, which can involve manual formatting and cleaning of the data using Excel, Microsoft Access, and/or scripted routines.

How does a user acquire / access data?
Depending on the licensing restrictions, a user may download data through the public-facing side of the web portal or request data from the dataset owner.

What user support services are available (both for depositing and accessing/using data)?
Phone and/or email support is available as needed. In-person training for both data import and data access may be offered depending on the circumstances. For priority projects or datasets, FEMC staff will help users deposit or access data they need.

How does the resource curate data at the time of deposit?
FEMC relies on the user to accurately create the metadata records, and relies on the University of Vermont information technology infrastructure to provide physical and digital backups of the data. FEMC staff conduct periodic reviews of metadata for completeness.

Technical characteristics and policies

Software platform description, incl. data search and access API(s):
JavaScript/JQuery/Bootstrap front end with PHP/MySQL backend. Limited search API functionality has been developed as a proof-of-concept for one partner. We provide a limited API for our DataONE slender node to access and process metadata objects using Python.

Service reliability (including recent uptime statistics, frequency of hardware refresh, if known):
We do not have access to this information, but could pursue it from UVM if needed.

Preservation reliability (including replication/backup, integrity checks, format migration, disaster planning):

FEMC relies on UVM Enterprise Technology Services (ETS) for the server infrastructure supporting our operation. This includes backup services, upgrade services and physical redundancy. ETS performs nightly backups of file and database servers, and maintains physically disconnected redundancy for these systems. We are not familiar with any integrity check routines that they perform, but can pursue this if required. By utilizing all open source software throughout our stack, we believe we're mitigating risks from future changes in format.

User authentication technology (incl. level of create/modify/delete access by users):
We have implemented the Ion Auth library for the CodeIgniter framework to manage authentication. Ion Auth provides password management, session tracking and lock-out methods. Once authenticated, there are several levels of user permission based on authentication and roles. Permissions are set on both project and dataset levels, where a project is a collection of datasets and related data. Project or dataset managers can make create, modify, and delete data and adjust user roles and permissions on those projects and datasets. Dataset editors can add, modify, or delete data, but not adjust user information. Those with read-only access are the most limited. System administrators are the only ones who can create new accounts, approve requests for accounts, or make changes to projects that they do not own.

Data identifier system and data citation policy, if available:
FEMC will soon implement the minting of DOIs through EZID. A recommended data citation string will be generated dynamically as the necessary metadata fields are populated. Data citation policy is set by the type of Creative Commons license chosen for the data, but is not enforced or monitored directly by FEMC.

Metadata standards (including provenance):
The FEMC adheres to the Ecological Metadata Language (EML) standards for documenting metadata. FEMC utilizes Open Geospatial Consortium standards for spatial data stored in the database. FEMC implements dataset versioning using the Research Data Alliance's Dynamic Data Citation recommendation, and implements data typing using the RDA's Data Types Registry recommendation. FEMC maintains a changelog for edits to dataset and project metadata, but does not currently display this to users on the site. Beyond dataset versioning and the changelog, we do not implement any direct provenance tracking at this time.

Capacity/services to DataONE

At what functional tier will you initially be operating? (see http://bit.ly/MNFactSheet for definitions)

- [x] **Tier 1: Read only, public content**
- [ ] **Tier 2: Read only with access control**
- [ ] **Tier 3: Read/write using client tools**
- [ ] **Tier 4: Able to operate as a replication target**

If you can host data from other member nodes, what storage capacity is available?
N/A

Can you provide computing capacity to the broader network? If so, please describe.
No.

Other Services

What other services or resources (such as expertise, software development capacity, educational/training resources, or software tools) can be provided of benefit to the broader network?
To be assessed as development evolves and time permits. FEMC staff are more than willing to be available to the broader network for guidance in implementation or to demonstrate tools built off the DataONE and FEMC infrastructure.