

Federation of Earth Science Information Partners

Data Management Training Clearinghouse

January 2016

Version 1.0

Contact

David Bassendine

Owner and Web Architect

david@bluedotlab.org

678-765-9566

Contents

[Summary](#)

[Work](#)

[1. Preparatory](#)

[i. Setup](#)

[2. Discovery](#)

[i. Information Architecture](#)

[ii. User Roles and Stories](#)

[iii. Wireframing](#)

[3. Development](#)

[i. Set up ESIP microsite](#)

[ii. Store and display Learning Resources](#)

[iii. Search and Listings](#)

[iv. Homepage](#)

[v. Crowdsourc Learning Resources](#)

[vi. Moderate crowdsourced Learning Resources](#)

[vii. Automated link checking](#)

[viii. Deployment](#)

[4. Ongoing](#)

[i. Project Management](#)

[ii. Contingency](#)

[Timeline](#)

[Estimate](#)

Summary

Prototype a Clearinghouse that pulls together learning resources for data management training from disparate sources. Direct visitors towards key topics from the data management lifecycle, or allow them to filter by facets or run freeform keyword searches. Allow visitors to suggest resources to include, then notify editors to check, adjust and publish them.

Jump-start development and ensure long-term sustainability by building the clearinghouse within the Federation for Earth Science Information Partners' (ESIP's) existing multi-domain Drupal environment at esipfed.org.

Work

Below, we list the work to be undertaken, including approaches and estimates, broken down into three main sections:

- **Preparation** - necessary setup to start Development work
- **Discovery** - rapid scoping to establish overall patterns
- **Development** - a series of self-contained work units, prioritized by you, then developed in priority order and deployed together at the end of the first phase
- **Ongoing** - support work not linked to any individual piece of work

We treat pieces of work as one of two types:

- a **Feature** is planned ahead in a discovery process that sets out a clear specification and estimate with an upper cap on time spent - the work is amenable to and benefits from forward planning
- a **Task** is best handled iteratively without a formal discovery process or upper cap on time spent - the work may be small, simple or hard to plan by its nature

The initial estimates provided here are our best effort based on preliminary information from initial discussions. As individual pieces of work are scoped out, we can narrow the uncertainties and provide more formal estimates with an upper cap, if the work is amenable to it (that is, for Features). We think it likely these initial estimates will be met but cannot absolutely guarantee it.

1. Preparatory

i. Setup

Allows the Project Manager and Developers to get the project ready for development. We're already set up to work on the ESIP sites, so we'll only need to set up our Project Management and Testing systems for this project.

Covers:

- Project Management setup
- Testing spreadsheet setup

Estimate: 2 hours

2. Discovery

A fast discovery process designed to establish overall patterns for the application, without delving into too much detail - which we leave to scoping and prototyping during development of each smaller feature. By keeping a small initial core, then layering on features, we:

- Respond to changes bubbling up during the project that can't be anticipated
- Deliver a usable application earlier in the project cycle
- Allow you to re-prioritize features on the fly, if need be

i. Information Architecture

Outline the structure of the application, in terms of page types, their likely pages and how they are linked together to form an overall site map. Based on those relationships, define URL patterns and the hierarchical structure of menus.

To keep this prototype as simple, straightforward and usable as possible, within initial budget constraints, we recommend the following broad structure, which informs the approach of the rest of the work listed here:

- Search/Listing Interface - single unified learning resource listing, filtered by topic and keyword search, showing key metadata for each one. Using faceted search, we can provide freeform discovery alongside directed browsing in a single interface, using a single technology stack and architecture.

- Homepage - adjusts the main search/listing interface to introduce the tool, provide guidance, and show learning resource highlights. By opening with the main interface, we bring visitors directly to the learning resources they are looking for, reduce navigational complexity and offer guidance whilst visitors are already *in* the main interface
- Learning Resource - dedicated page and URL for each learning resource, serving as a hub showing its external location, full description, organization and the rest of its metadata, and in future, community annotations including ratings and comments. Provides space for an in-depth look at each learning resource, making each one an SEO first class citizen for easy discovery by search services (eg. Google).

Note that topics don't have a dedicated identity in this structure - they are one facet amongst many, albeit a prominent one, within the main search interface.

Covers:

- Overall structure
- Menus
- URL paths
- Identify key content pages

Estimate: 2 - 3 hours

ii. User Roles and Stories

Define key user roles and describe how each will use the application.

From initial discussions, there are four user roles:

- Anonymous - can search and browse the Clearinghouse
- Submitter - above + can submit new learning resources (for moderation by editors)
- Editor / Peer Reviewer - above + submission notification, editing and publication
- Administrator - above + configure the overall structure of the application

Covers:

- Define user roles
- Outline key user stories
- Update testing spreadsheet

Estimate: 1 - 2 hours

iii. Wireframing

Identify functional page elements that are needed for each page type. Then position those elements in a responsive ayout, with arrangements for mobile, tablet and desktop.

To streamline this prototype, we'll reuse existing layouts already available within the ESIP environment, working within their constraints during this project.

Page types:

- Homepage and Search (combined interface)
- Learning Resource (based on ESIP standard)

Covers:

- Page descriptions (x2)
- Wireframe layouts (x2) (each responsive)

Estimate: 3 - 6 hours

3. Development

i. Set up ESIP microsite

To jump start development and improve the sustainability of long-term hosting and maintenance costs, we'll prototype the Clearinghouse within ESIP's existing multi-domain Drupal environment at esipfed.org.

The Clearinghouse would be a new subsite within the ESIP family, accessible directly or via the "Explore" menu from other ESIP subsites. It would inherit ESIP's look and feel and responsive layouts, but define its own identity through a distinct navigation bar title and having its own subdomain of esipfed.org (such as datamanagement.esipfed.org). The layouts and content shown in the main content area remain flexible, as does the navigation menu itself.

We'll use the Domain module to separate content, search indexing and layouts for the clearinghouse from existing esipfed.org subsites. This approach is already being used successfully to provide 3 different esipfed.org subsites (commons.esipfed.org, testbed.esipfed.org and esipfed.org). By sharing a common backend, development and maintenance costs are vastly reduced, because each site can inherit the codebase and Drupal architecture of the others - only overriding defaults where absolutely necessary.

We'll inherit a rich architecture that will jump start development and allow us focus on work that really adds value. From day one, we'll benefit from ESIP's existing:

- Hosting environment, optimized for Drupal with scalable Varnish caching, Solr

search indexing and New Relic monitoring (Pantheon, see <http://getpantheon.com>)

- Drupal build (core, content types, search, and other contributed modules)
- Bootstrap frontend framework (streamlines design, styling and testing)

Covers:

- Scoping
- Configure domain module
- Configure menus
- Configure blocks & context
- DNS Setup

Estimate: 2 hours

ii. Store and display Learning Resources

Store learning resources by defining their field schema, including metadata fields. Display each learning resource at a unique URL, with layouts for mobile, tablet and desktop.

As each learning resource inside the Clearinghouse points to an external resource at a particular URL, we only need one field schema. The type of media used (or other classifications) can be defined by taxonomies rather than differentiated schemas.

We'll set up a single new content type in Drupal, then add fields to define its schema. By enabling Drupal's built-in revisioning system, we can log when a learning resource is created or modified. Every revision is timestamped and tied to a user account, and differences between revisions can be reviewed by editors or peer reviewers in the admin interface.

We'll then flexibly map fields to responsive frontend layouts, by mapping fields to blocks, then assigning those blocks to regions within our Bootstrap theme. We can also adjust layouts or display styles for specific devices using Bootstrap classes. By inheriting the ESIP look and feel from the existing theme, we save time on basic design and styling and can focus on clearinghouse-specific highlights and adding new features. For example, we will brand each Learning Resource with the organization who provided it.

Covers:

- Scoping
- Content type x1 (bundle + fields configuration)
- Display configuration (fields view mode x1)
- Page layout configuration (display suite, block configuration)
- URL Path settings
- Styling

- Testing

Estimate: 7 - 11 hours

iii. Search and Listings

Index learning resources so they can easily be discovered by keyword search or browsing by different facets (eg. steps of the USGS Science Data Lifecycle, media type, or one or more metadata fields) in an intuitive interface.

For browsing by topic - it's more straightforward to use a hierarchical taxonomy as a facet within the search interface, than set up two separate architectures for topic browsing and searching. Essentially, the topic becomes one of many facets by which visitors can browse the Clearinghouse, albeit one that's given prime visibility. One drawback of this approach, is that it then becomes harder to display explanatory text about each topic above its listing. We think this is important to provide much needed context for each topic, so we've included some time to review approaches to accomplish this within the Search interface.

For the listing - we recommend initially displaying only the title and type of each resource, so visitors can quickly scan the results list for something they're interested in. When they find it, they would tap or click to expand it, review some more details, then - if it stills looks good - click on the button to see the full entry.

We could use the following examples of similar faceted search and listing interfaces as a starting point for the interface we design here:

- ESIP Semantic Web Syndicator (<http://esip.semanticweb.syndicator.scienceondrupal.org>, a tool to aggregate community content via RSS, mash it up and reshare) - a content listing with expandable teasers, and a simple filter mechanism
- C-DEBI Data Portal (<http://cdp.darkenergybiosphere.org>, archive of publications and datasets for the Center for Dark Energy Biosphere Investigations) - faceted search interface and content listing, showing metadata for each publication

We'll leverage ESIP's existing Apache Solr server and Drupal integration. The Learning Resource content will be indexed alongside ESIP's other content, but filtered out when using Views to query Solr and build listings for the user interface.

Covers:

- Scoping
- Design (small adjustments, type icon set)
- Search, facets and sort set up (drupal solr & block configuration)

- Facet specific content (descriptions)
- Listing (view, teaser view mode, read more button, show/hide on title click)
- Styling (apply bootstrap classes)
- Testing

Estimate: 14 - 22 hours

iv. Homepage

Welcome visitors by introducing the clearinghouse, giving them some tips to get started and highlighting particular learning resources in the listing area. Otherwise, keep the interface the same as the faceted search interface.

We recommend this approach to keep the application simple, straightforward and easy to use - introducing the search interface from the start. This also avoids adding unnecessary scope to the first phase of the project.

We'll configure some blocks to appear (or expand) when on the homepage, then disappear (or contract) out of the way when the user starts searching or browsing the catalog. By setting default facet selections, we can also highlight particular resources on the homepage when visitors first arrive.

Covers:

- Scoping
- Design (adjust bootstrap defaults)
- Overall architecture (context and blocks)
- Welcome and mission
- Search facets for homepage
- Styling
- Testing

Estimate: 2 - 3 hours

v. Crowdsourcing Learning Resources

Enable and encourage visitors to submit new learning resources to the Clearinghouse.

We'll highlight this capability by showing a prominent button that takes visitors straight to the submission form. Once there, we'll lower their barrier to entry by hiding some fields, using friendly form widgets and showing guidance for each field. To promote more accurate metadata entry, we'll show contextual descriptions for each tag visitors could add to each resource, clarifying what they really mean.

To implement this, we'll open up access to Drupal's existing "add content" form for the Learning Resource content type we created in *2ii Store and display Learning Resources*. Setting per-field help text can provide guidance for each field, and an inline tag glossary can be provided using the Term Reference Filter by Views module.

Covers:

- Scoping
- Install and configure field_permissions
- User friendly field widgets
- Inline field help text
- Inline descriptions of tags
- Button to add content
- Styling
- Testing

Estimate: 3 - 6 hours

vi. Moderate crowdsourced Learning Resources

Working with the Clearinghouse team, we'll outline a workflow process that allows editors and peer reviewers to evaluate and edit submitted Learning Resources effectively. At a minimum, we'll notify moderators when new learning resources are submitted, and provide a moderation dashboard that makes it easy for peer-reviewers or editors to check and adjust Learning Resources, before finally approving them for publication.

We'll use the Workflow module to set up appropriate workflow states, along with Rules to adjust publication status and send email notifications when necessary. A View will provide a moderation queue in Drupal's admin interface that moderators can check to see what's waiting for them to work on.

Covers:

- Scoping
- Install workflow modules
- Workflow states set up (workflow module)
- Publication status setup (rules module)
- Notifications setup (rules)
- Moderation queue (view x1)
- Styling
- Testing

Estimate: 6 - 12 hours

vii. Automated link checking

Automate checking for broken URLs in Learning Resources. This will help the Clearinghouse team fix broken links or hide deprecated resources in a timely manner - as well as dramatically reducing their moderation workload.

We'll use the Link Checker module to add this automation very quickly in a way that's well supported by the Drupal community. One potential drawback to this approach, though, is that moderators must check a report in the admin interface to be aware of broken links. If email notification is needed, we'd need to integrate it ourselves - this is something we could look at, but is not included here.

Covers:

- Scoping
- Install linkchecker + httpprl
- Set up linkchecker
- Testing

Estimate: 1 hour

viii. Deployment

After having developed all the features listed above, tested them on an independent staging environment, and gained approval from the Clearinghouse team - we'll deploy these together to the live ESIP environment.

Just after deployment, the Clearinghouse team will populate the Clearinghouse with its initial live content, drawing from the ESIP Data Stewardship Committee's Survey, including USGS Data Management and DataONE education modules.

Covers:

- Planning
- Configuration on live
- Push to live, DNS change
- Testing

Estimate: 2 - 3 hours

4. Ongoing

i. Project Management

Covers:

- Meetings or liaising with other stakeholders
- Weekly progress reports
- Ticketing and work organization

Estimate: 5 - 8 hours (10% of subtotal)

ii. Contingency

Allows for inherent uncertainties that may emerge during scoping or development of listed work, or for additional pieces of work that may be needed.

Estimate: 3 - 4 hours (5% of subtotal)

Timeline

We'll be able to provide around 15 hours per week developer time for this project.

We'll set up the new subsite and develop its new features on top of the existing ESIP site build - but keep the work separate using version control branches for code, and a separate staging environment for testing and demonstration. Once all the features are complete, tested using sample content and approved by your team, we'll deploy to the ESIP live environment. Then, we can populate the Clearinghouse with real content, before its final launch.

We anticipate the project would take **6 calendar weeks** to complete and deploy. Here are interim milestones, relative to the start week (#1) - which is the first full week we have available once the contract is finalized.

Work items	Completion Wk
1. Preparation	1

2i. Information Architecture	1
2ii. User Roles and Stories	1
2iii. Wireframing	1
3i. Set up ESIP microsite	1 - 2
3ii. Learning Resource	2
3iii. Search and Listings	3 - 4
3iv. Homepage	3 - 4
3v. Crowdsourced Learning Resources	3 - 4
3vi. Moderate crowdsourced Learning Resources	3 - 5
3vii. Automated link checking	3 - 5
3viii. Deployment	4 - 6

This timeline assumes that:

1. actual development times fall within our broad estimates - likely, but times can change if more work emerges during scoping or development)
2. no additional work is needed before deployment - likely, but new work can emerge during testing

Estimate

As noted above, times given are estimates and may need to be revised based on scoping outcomes for each individual piece of work. Estimates are in units of hours and include an allowance for:

- Discovery - discussing, scoping and specifying the work to be done
- Development - carrying out the work itself, and testing it
- Documentation - outlining the architecture and administration of the new feature

Work items	Type	Estimate (h)
------------	------	--------------

1. Preparation	Task	2
2i. Information Architecture	Task	2 - 3
2ii. User Roles and Stories	Task	1 - 2
2iii. Wireframing	Task	3 - 6
3i. Set up ESIP microsite	Task	2
3ii. Learning Resource	Feature	7 - 11
3iii. Search and Listings	Feature	14 - 22
3iv. Homepage	Feature	2 - 3
3v. Crowdsourced Learning Resources	Feature	3 - 6
3vi. Moderate crowdsourced Learning Resources	Feature	6 - 12
3vii. Automated link checking	Feature	1
3viii. Deployment	Task	2 - 3
Subtotal		45 - 73
4i. Project Management (required, 10% of subtotal)	Task	5 - 8
4ii. Contingency (required, 5% of subtotal)	Task	3 - 4
Total (hours)		53 - 85
Total (\$)		4,770 - 7,650

The work would be billed on a time and materials basis at a rate of \$90/hr, billed monthly, up to a spending limit defined by the overall estimate, whichever is lower. The length of time taken for each Feature may vary after Discovery is complete, so we cannot absolutely guarantee the amount of work to be completed within the overall limit. We do, however, think it likely these estimates are realistic and will provide sufficient resources to meet them within the agreed timeline. We will provide regular feedback on progress and time spent, allowing you to reprioritize as necessary to ensure all your important pieces of work are completed.