



Interoperability Using Lightweight Metadata Standards: Service & Data Casting, OpenSearch, & OPM Provenance

B. Wilson¹, G. Manion¹, H. Hua¹, R. Ramachandran², A. Kulkarni², M. Maskey², K. Keiser² and S. Graves²

¹Jet Propulsion Laboratory, ²University of Alabama Huntsville

Summary

The Infocasting project has three main goals:

1. Define and evangelize micro-formats (metadata standards) so that providers can easily advertise their web services, datasets, and topical geophysical events by adding structured information to broadcast feeds (Atom or RSS XML);
2. Develop authoring tools so that anyone can easily author such service advertisements, data casts, and event descriptions;
3. Provide a one-stop, Google-like search box in the browser that allows discovery of service, data and event casts visible on the web, and services & data registered in GCMD, ECHO and other repositories, like the GEOSS registry.

Version 2.0 of the Service Casting standard has been finalized and beta testers are using it. A “smart” authoring tool makes it easy to create and publish “scasts”. Authors can publish the service advertisement on the ServiceCommons web site or export it to publish on their own web site. We are currently aggregating service casts from early adopters in the ESIP Federation, and wider evangelism will follow.

Many kinds of metadata can be “cast” in Atom feeds: service advertisements, collection metadata, the existence of each data granule as it is produced, and hybrid feeds bundling information about topical geophysical events. An event cast consists of a short micro-article with a key figure depicting the event, along with a bundle of relevant datasets, services, canned OpenSearch queries, and even analysis workflow documents. The first version of the Event Casting format has been defined and we are currently prototyping the event cast authoring tool. Building on the casting formats, we have developed multiple interfaces to search & browse the metadata, including: faceted navigation of the ECHO collections; federated OpenSearch for collections & granules across ECHO, MIRADOR, NSIDC, and other back-ends; and search/browse of service and event casts.

The research described here was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Definitions

- **Kinds of casts:**
 - **Service cast (scast)** – Atom feed, advertise bundle of services
 - **Dataset** or collection cast – Atom feed, collection level metadata
 - **Datacasting** (A. Bingham, JPL) – RSS feed, announce data granules as they are produced, links to data & browse image
 - **Event cast** – Hybrid Atom feed of event features (micro-article) and relevant data & services
- **OpenSearch Protocol**
 - Submit query for services, datasets or granules as one-line URL
 - Result set returned in pages as a feed, using scast, collection cast and datacast formats
- **Synergy between Casting and OpenSearch**
 - When publishing metadata in feed or in OpenSearch response, same format used for both cases.

Who Benefits?

- **Data producers & service providers**
 - Advertise their collections, granules, and services
 - **Ownership:** Publish metadata on their own web sites
 - **Control:** Change & re-publish at will
 - **Flexible:** Mix of standard & proprietary fields
- **Platform for Innovation (“middleware” providers)**
 - **Publish & Subscribe:** “push” metadata updates
 - **Scalable:** Like Google, can crawl & aggregate
 - Many aggregators can provide search and add value
- **Scientists and the public**
 - **Open metadata** → Discovery capabilities not limited to what latest registry provides. Marketplace of competing interfaces.
 - Rich “faceted” navigation of datasets & services
 - Smart ontology-enhanced search (Noesis 2.0)
 - Event casts – subscribe to a curated data/feature album that grows as analysis proceeds and items are added to the feed.
- **Disaster Response Scenario**
 - **Disaster or geophysical event happens**
 - First images or features emerge
 - **Scientist hears of event and begins analysis**
 - Finds datasets & services using “one-stop search box”
 - Discovers OGC/WMS server that returns LANDSAT images
 - Performs OpenSearch at ECHO or Mirador for MODIS granules at appropriate time and spatial region
 - Creates analysis workflow to bring together several data sources
 - Generates new image (or plot) of the phenomena
 - **Scientist assembles & publishes event cast**
 - Short two-paragraph description of his generated image
 - Curated data album of relevant data granules & images
 - Relevant services (the WMS & OpenSearch servers)
 - Adds SciFlo or BPEL workflow document
 - Publishes the bundle as a subscribe-able feed (the cast)
 - Sends URL of feed to other scientists
 - **Scientists hit URL in browser, which displays rich GUI using the event data**

Service Cast Browser

The screenshot shows a web browser interface for service casts. On the left, there's a 'My Feeds' sidebar with various service icons like 'SciFlo Data Query/Access Services (JPL)', 'RGIS Base Imagery Service', and 'ADAM Services: Data mining and image process'. The main area displays a list of feeds, with one selected: 'GeoRegionQuery' by Brian Wilson. Below the list, there's an 'Example Call' section showing a GeoRegionQuery URL with parameters like 'dataset=AIRS', 'level=L2', 'version=none', etc.

Peer-Reviewed Micro-Articles

The screenshot shows the 'Journal of Earth Science Phenomena' website. It features a search bar at the top and a navigation menu. The main content area displays a list of micro-articles, including 'Smoke from an Oil Refinery Fire in Puerto Rico' by Ramachandran, et al. and 'Australian Dust Storm – A Satellite View' by Pawan Gupta, Falguni Patadia, Sundar A. Christopher. Each article includes a title, date, and a small thumbnail image.

Service Cast Authoring Tools

The screenshot shows the 'Service Cast Authoring Tools' interface. It includes a 'Service Commons' header, a 'Create a Service Cast' button, and a form for entering service details. The form fields include 'Service Title', 'Service Feed ID', 'Service Summary', 'Service Semantics', 'Service Protocol', and 'Service Keywords'. There are also sections for 'Links' and 'Additional Tags'. A 'Save Service Cast' button is at the bottom right.

Search & Browse Interfaces

This block contains two screenshots. The left one is the 'Federated OpenSearch Interface', showing a search results page with various filters and a list of search results. The right one is the 'Graph of Production Provenance (OPM)', which is a network diagram showing relationships between different data sources and services, with nodes representing services and edges representing data flows.

Faceted navigation of ECHO Collections & Services

The screenshot shows the 'Faceted navigation of ECHO Collections & Services' interface. It features a search bar and a list of search results. The results are organized into faceted categories, such as 'Processing Center', 'Platform', and 'Instrument'. Each category has a list of items with their respective details and links.