

# BTAA GDP Accession Guidelines for ArcGIS Open Data Portals

Version 1.0 - April 18, 2018

---

## OVERVIEW

This document describes the BTAA GDP Accession Guidelines for ArcGIS Open Data Portals, including eligible sites and records, harvest schedules, and remediation work. This policy may change at any time in response to updates to the ArcGIS Open Data Portal platform and/or the BTAA GDP harvesting and remediation processes.

## ELIGIBLE SITES

**Policy:** Any ArcGIS Open Data Portal (Arc-ODP) that serves **public** geospatial data is eligible for inclusion in the BTAA GDP Geoportal (“the Geoportal”). However, preference is given to portals that are hosting original layers, not federated portals that aggregate from other sites. Task Force members are responsible for finding and submitting Arc-ODPs for inclusion in the Geoportal. Each Arc-ODP will be assigned a Provenance value to the university that submitted it or is closest geographically to the site.

**Explanation:** In order to avoid duplication, records that appear in multiple Arc-ODPs should only be accessioned from one instance. This also helps to avoid harvesting records that may be out of date or not yet aggregated within federated portals. Although the technical workers at the University of Minnesota will be performing the metadata processing, the Task Force members are expected to periodically monitor their records and make suggestions for edits or additions.

## ELIGIBLE RECORDS

**Policy:** The only records that will be harvested from Arc-ODPs are **Esri REST Services** of the type *Map Service*, *Feature Service*, or *Image Service*. This is further restricted to only items that are harvestable through the DCAT API. By default, the following records types will **not** be accessioned on a regular basis:

- Web applications
- Nonspatial data, tabular data, PDFs
- Single records that describe many different associated files to download, such as imagery services with a vast number of sublayers

**Explanation:** Arc-ODPs are structured to automatically create item records from submitted Esri REST services. However, Arc-ODP administrators are able to manually add records for other types of resources, such as external websites or documents. These may not be spatial datasets and may not have consistently formed metadata or access links, which impedes automated accessioning. If these types of resources are approved by the Metadata Coordinator, they may be processed separately from the regular accessions.

## QUERY & HARVEST FREQUENCY

**Policy:** The Arc-ODPs included in the Geoportal will be **queried monthly** to check for deleted and new items. The results of this query will be logged. Deleted items will be removed from the geoportal immediately. New records from the Arc-ODPs will be accessioned and processed within two months of harvesting the metadata.

**Explanation:** Removing broken links is a priority for maintaining a positive user experience. However, accessioning and processing records requires remediation work that necessitates a variable time frame.

## REMEDIATION WORK

The records will be processed by the Metadata Coordinator and available UMN Graduate Research Assistants. The following metadata remediation steps will be undertaken:

1. A Python script will be run to harvest metadata from the DCAT API. This will provide the following elements for each record:
  - Identifier
  - Title
  - Description
  - Date Issued
  - Date Modified
  - Bounding Box
  - Publisher
  - Keywords
  - Landing Page
  - Web Service link
2. The metadata will be batch augmented with Administrative template values in the following elements:
  - Collection
  - Rights
  - Type
  - Format
  - Provenance
  - Language
  - Centroid (derived from Bounding Box)
  - Download Link (created from Landing Page)
  - Tag (web service type)
  - Thumbnail link (derived from server or Arc-ODP page)
3. The metadata will be manually augmented with descriptive values for the following elements:
  - Subject (at least one ISO topic category)
  - Geometry type (Vector or Raster)
  - Spatial Coverage (place names written out to the nation level: “Minneapolis, Minnesota, United States”)
  - Temporal Coverage (dates included in the title or description)
  - Title (add place names, expand acronyms, and move dates to the end of the string)
  - Description (remove html and non-UTF8 characters)
  - Creator (if available)
  - Solr Year (integer value based on temporal coverage or published date)
4. The metadata will **not** be fully remediated for the following cases:
  - Missing bounding box coordinates (0.00 values) will be defaulted to the bounding box of administrative level of the Arc-ODP or the record will be omitted.
  - Missing or incomplete descriptions will be left alone or omitted from the record
  - Individual records that require additional research in order to make the metadata record compliant, such as missing required elements or non-functioning links, will be omitted.

## STANDARDS METADATA

**Policy:** Creating or linking to standards based metadata files for Arc-ODPs is out of scope at this time.

**Explanation:** If metadata is enabled for an Arc-ODP, it will be available as ArcGIS Metadata Format 1.0 in XML, which is not a schema that GeoBlacklight can display. The metadata may also be available as FGDC or ISO HTML pages, but these types of links are not part of the current GeoBlacklight schema. Further, very few Arc-ODPs are taking advantage of this feature at this time.