

Downloading species observations from the Ocean Biogeographic Information System (OBIS) with the DiGIR protocol

THIS EXAMPLE IS UNDER CONSTRUCTION

This example shows how invoke MGET tools from ArcGIS to download species observation records from OBIS and save them as a point feature class. This example assumes you have basic familiarity with ArcGIS.



What are OBIS and DiGIR?

The [Ocean Biogeographic Information System \(OBIS\)](#) is a strategic alliance of people and organizations sharing a vision to make marine biogeographic data, from all over the world, freely available over the World Wide Web. OBIS maintains a database of species observations contributed by member organizations and individuals. At the time of this writing, the database contained over 16 million records. Most records include a latitude, longitude, and date, making them suitable for geospatial analysis.

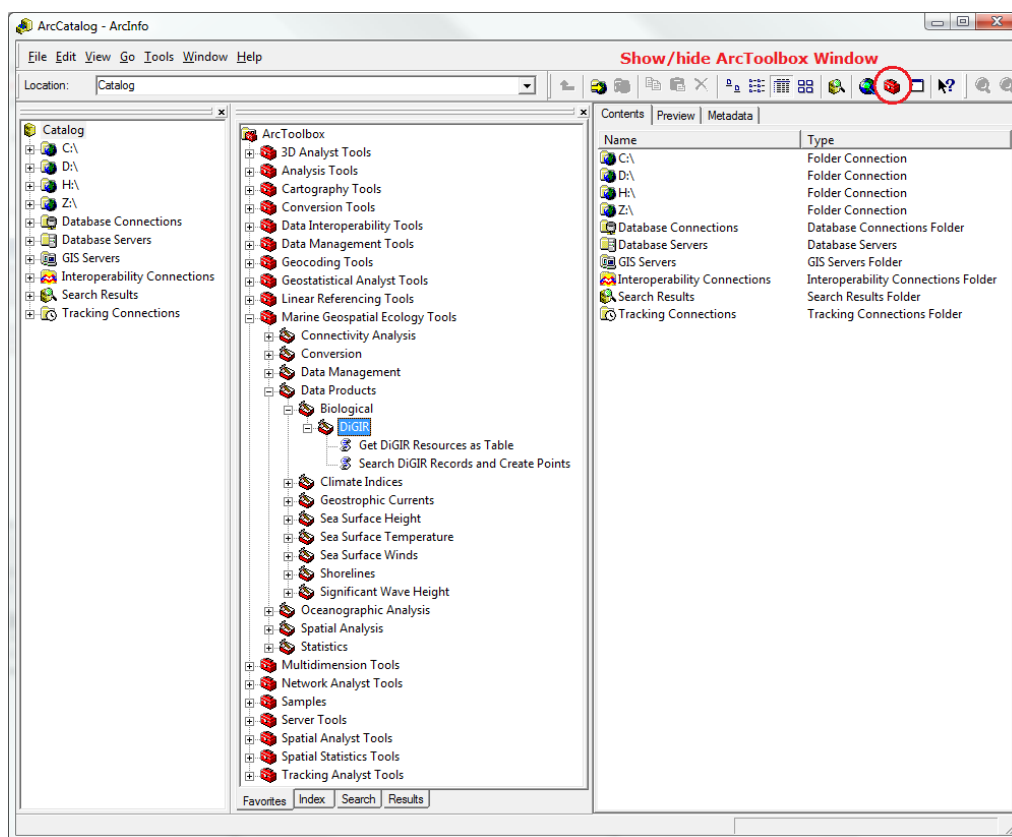
Records may be extracted from the OBIS database with the [Distributed Generic Information Retrieval \(DiGIR\)](#) protocol. This protocol is in widespread use throughout the bioinformatics community. OBIS uses it to collect records from contributing organizations such as [OBIS-SEAMAP](#), a database of marine mammal, seabird, and sea turtle records maintained here at Duke University. The [Global Biodiversity Information Facility \(GBIF\)](#) uses it to collect records from OBIS into a larger biogeographic database for all taxa. At the time of this writing, there were dozens of servers that implemented the DiGIR protocol. The [Big Dig](#) website maintained a list of DiGIR servers. Big Dig stopped operating in February 2008 but many of the servers listed there are still operational and can be queried with DiGIR.

Although DiGIR is widely used, it is no longer under development. A new protocol, [TAPIR](#), may eventually replace it. And although GBIF uses DiGIR to collect data from its many contributors, it exposes data through several [GBIF-specific web services](#). As far as we know, you cannot query GBIF with DiGIR.

Like many "web service" protocols, DiGIR is an XML-based stateless request/response protocol. For technical details, please see the documents and XML schemas on the [DiGIR home page](#).

Finding MGET's DiGIR tools in ArcGIS

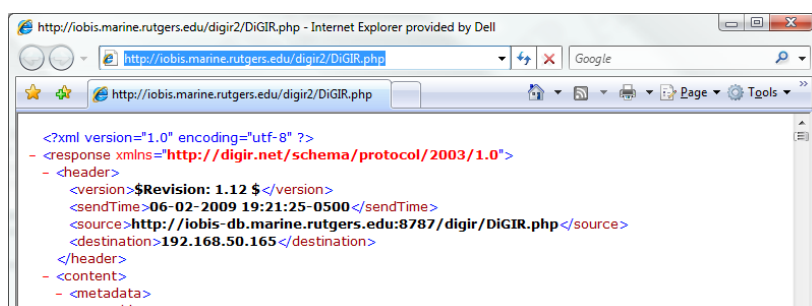
You must have MGET 0.7a10 or a later release to use the tools. They appear in the Marine Geospatial Ecology Tools node in the ArcToolbox window. If you do not see the ArcToolbox window, click the **Show/Hide? ArcToolbox Window** button on the toolbar.



DiGIR server URLs

Before you can use the MGET tools, you must obtain the URL for the server you wish to query. The best way to obtain the URL is to contact the server operator. The [Big Dig](#) website contains a list of several hundred URLs for servers that were active as of February 2008. (That website uses the formal DiGIR term *provider* rather than *server*. The terms are synonymous.) At the time of this writing, the URL for the OBIS server was <http://iobis.marine.rutgers.edu/digir2/DiGIR.php>. The URL for the OBIS-SEAMAP server maintained by Duke was <http://seamap.env.duke.edu/digir/DiGIR.php>.

Before you attempt to use the MGET tools, it is a good idea to enter the server URL into a web browser to make sure the server is responding. Some servers may take several minutes to respond if they have not been accessed by anyone for a while. Eventually you should receive some XML or be prompted to save a document that contains XML. For example, the OBIS server returns a response that looks like this in Internet Explorer:



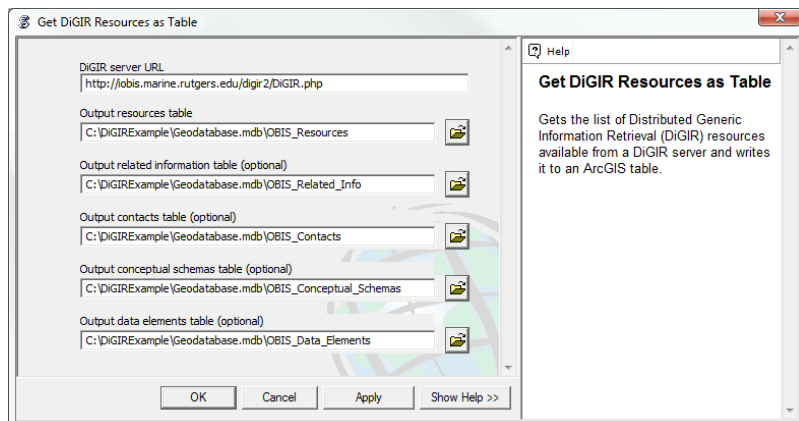


If you get a response similar to this, it is likely that the server is functional and the MGET tools will work. If you don't, something is wrong and the MGET tools will probably fail. You should follow up with the server operator to resolve any problems.

Discovering the resources available from a DiGIR server

Each DiGIR server hosts a set of *resources*. A resource is a collection of related species observation records. Once you obtain a URL to a server, it is possible to immediately start downloading these records, but it is usually worthwhile to first understand what resources are available from the server and review some metadata about them. MGET's **Get DiGIR Resources as Table** tool assists with this process.

The input parameters to the tool are the URL to the server and the paths and names of five tables of metadata:



Important: We suggest you store the output tables from this tool in a geodatabase rather than DBF files. Text fields in DBF files are limited to 254 characters, and several of the fields of this table will usually exceed this limitation.

The most important output is the resources table:

Code	NumRecords	LastUpdate	Name	Abstract	Keywords	Citation	UseRestr	RecordID	RecBasis
EMAP	39882	8/23/2007 2:22:30 PM	EPA'S EMAP Database			Request for Acknowledgm	Request for Acknowledgm	Unique identifier for this resou	voucher - can be voucher o
HMAP	242384		HMAP-History of Marine Ani			http://www.hmapcoml.org/		hmap	
NODC	1275382	8/23/2007 2:22:30 PM	NODC WOD01 Plankton Data	Coming soon	Coming soon	M.E. Conkright, J.I. Antonov,	http://www.nodc.noaa.gov	Unique identifier for this resou	voucher - can be voucher o
SEAMOUNTS	7318	8/23/2007 2:22:30 PM	SeamountsOnline (seamoun	SeamountsOnline holds d	seamount, guyot, specie	Stocks, K. 2003. Seamount	Users must agree to cite th	Unique identifier for this resou	voucher - can be voucher o
SOC	92851	8/23/2007 2:22:30 PM	Southampton Oceanography	Description of the content	Some key words (comm	How you want your data ci	Restrictions on use of the	Unique identifier for this resou	voucher - can be voucher o
ZOOGENE	114	8/23/2007 2:22:30 PM	ZooGene A DNA Sequence	An international partnershi				Unique identifier for this resou	voucher - can be voucher o
BATS	635	8/23/2007 2:22:30 PM	Bermuda Atlantic Time-serie	Abundance and taxonomi	zooplankton, Bermuda,	Steinberg, D.K. and L.P. Ma		Unique identifier for this resou	voucher - can be voucher o
SERTC	2050	8/23/2007 2:22:30 PM	The SERTC Invertebrate Dat	The Southeastern Region				Unique identifier for this resou	voucher - can be voucher o
GISA	1549	8/23/2007 2:22:30 PM	Natural Geography In Shore	This dataset provides dat	Intertidal, Subtidal, Seagr	(how users should cite dat	No restrictions	Unique identifier for this resou	voucher - can be voucher o
cpr_data_sahfos1	721921	6/25/2004 4:20:27 PM	Continuous Plankton Record	The Sir Alister Hardy Foun			Restrictions on use of the	SAHFOS_CPR_PHYTOPLANK	
cpr_data_sahfos2	1374234	6/25/2004 4:20:27 PM	Continuous Plankton Record	The Sir Alister Hardy Foun			Restrictions on use of the	SAHFOS_CPR_ZOOPLANKTO	
fishbase	788855	6/3/2008	FishBase DiGIR Provider - P	FishBase is a large inform	FishBase, Fish, Fisherie	Froese, R. and D. Pauly. Ed	FishBase receives over 10	Unique identifier for this resou	voucher - can be voucher o
nbi	156864	11/8/2006 4:22:21 PM	NBI	NBI is a quantitative datab	benthic, inventory, NOA	For citation format please c	For restrictions please con	Unique identifier for this resou	voucher - can be voucher o
AK_observer	422150	8/28/2005	North Pacific Groundfish Ob			***Not yet provided***	None	Groundfish Observer 1993-20	observation
GuLFofMaineBIS	6155	4/19/2006 5:03:30 AM	Gulf of Maine Bottom Trawl	Description of the content	Northeast Fisheries Scienc	Bernice P. Bishop Museum	GBIF Data Use Agreement	CollectionCode	observation
BerniceP.BishopMuseum	8029	8/23/2007 8:35:50 AM	Bishop Museum Data (OBIS	Description of the content	Bernice P. Bishop Museum	MICROBIS	GBIF Data Use Agreement	CollectionCode	voucher
microbis	2584	8/23/2007 2:37:07 PM	MICROBIS database (CoML)	MICROBIS is a database o	Protista,Algae,Bacteria,e		All commercial use of imag	Unique identifier for this resou	voucher - can be voucher o
nemys	3616	8/23/2007 8:39:04 PM	Generic Taxonomic Databas	Nellys is a biological onlin		Deprez, T. (2000). Nellys,	Data are freely available th	NeMys	observation
obis	16207	8/23/2007 2:03:44 PM	Academy of Natural Science	The primary objective of t	Mollusca, Indo-Pacific, M	Rosenberg et al., 2002	The Academy of Natural S	ANSP Indo-Pacific Mollusc Dat	voucher
OBIS-SEAMAP	1033268	7/23/2006 3:38:56 AM	SEAMAP - marine mammals	The Ocean Biogeographic	marine mammal, bird, turt	Read, A. J. and P. Halpin. E	By accepting this document	OBIS-SEAMAP	
hexacorall	47117	8/23/2007 1:41:13 PM	Hexacorall Database	Description of the content	Some key words (comm	How you want your data ci	Restrictions on use of the	Unique identifier for this resou	voucher - can be voucher o
ECBMF	5650	6/9/2004 12:00:00 PM	Eastern Canada Benthic Ma	Measurements of biomass	Animal Taxonomy; Aqua	Database of Benthic Macro	Acknowledge the use of s	ECBMF	
ARC	125272	3/22/2006 4:53:23 PM	Atlantic Reference Centre (Summary: This is the Atla	Animal Taxonomy; Aqua	Atlantic Reference Centre	Please acknowledge the A	ARC	
EAISNSA	3437	12/22/2003 4:00:00 AM	Electronic Atlas of Ichthyopl	The goal of MIDI is to provi	Animal Taxonomy; Aqua	EAISNSA - An Electronic At	Acknowledge the use of s	EAISNSA	
GHMP	6351	2/5/2003 5:00:00 PM	Gwaii Haanas Marine Plants	The database covers the	Aquatic Habitat; Marine	Living marine legacy of Gw	Acknowledge the use of s	GHMP	
CMN	29877	10/31/2002	Canadian Museum of Nature	Summary: This digital data	Animal Taxonomy; Aqua	Canadian Museum of Natur	Data usage is governed by	CMN	
ACCDC	1365	5/31/2001	Atlantic Canada Conservatio	Summary: The Atlantic Ca	Community Structure; En	Atlantic Canada Conservati	Acknowledge the use of s	ACCDC	
NSM	579	7/29/1996	Nova Scotia Museum of Nat	This is the Nova Scotia Mu	Animal Taxonomy; Aqua	Nova Scotia Museum of Nat	Acknowledge the use of s	NSM	
BOFET	2381	2/5/2004 5:00:00 PM	Bay of Fundy Species List (Summary: The Bay of Fun	Huntsman Marine Scienc	Pohle, G., L. Van Guelpen,	Acknowledge the use of s	BOF	
MIDI	295	1/3/2004	Marine Invertebrate Diversi	The goal of MIDI is to provi	Animal Taxonomy; Aqua	Marine Invertebrate Diversi	Acknowledge the use of s	MIDI	
ECNASAP	466736	3/24/2006	ECNASAP - East Coast Nort	Summary: Fishery-indepe	Fisheries; Animal Taxon	East Coast North America S	Acknowledge the use of s	ecnasap	
RESOLUTE	3428	2/7/2005	Resolute Passage Copepod	This dataset includes data	Animal Taxonomy; Aqua	Copepod species compositi	Acknowledge the use of s	RESOLUTE	
gmbenthos	244	3/21/2005	Grand Manan Basin Benthos	This study investigates th	Marine Invertebrates;be	Grand Manan Basin - Deep	Not to be used for publicati	Grand Manan Basin Benthos	
ARCTIC80SBIONESS	9767	6/22/2005	Davis Strait and Baffin Bay	Summary: The purpose of	BIONESS	Lewis, M.K. and D. Smeot	Acknowledge the use of s	ARCTIC80SBIONESS	
MARWHALE	22783	5/15/2006 6:33:46 PM	Canada Maritimes Regional	This is an OBIS formatted	Earth Science > Oceans	K. Smedbol, J. Gale, L. Bajo	Acknowledge the use of s	marwhale	
POST	68510	2/22/2007 4:00:11 AM	Pacific Ocean Shelf Trackin	The Pacific Ocean Shelf T		Welch, D. W., G. Kristianso	Acknowledge the use of s	POST	

The important fields of this table are:

- Code** contains the server's abbreviation for the resource. When you use the **Search DiGIR Records and Create Points** tool, you can provide a list of codes to restrict the search to specific resources. Also, the points output by that tool have an attribute called **ResourceCode**, allowing you to determine which resource provided each point.
- NumRecords** shows the number of species occurrence records available from the resource and **LastUpdate** shows when they were last updated.
- Name**, **Abstract**, and **Keywords** describe the resource. This information is provided by the organization or individual that contributed the resource.
- Citation** gives instructions for citing the resource and **UseRestr** specifies any use restrictions imposed by the contributor. You are responsible for properly citing the resources you use and for observing these restrictions.

The remaining four tables are less important. The related information table lists the home page for each resource. The contacts table lists names, telephone

numbers, and email addresses of the people who contributed the resource. The conceptual schemas table lists the formal names of the technical documents that specify the fields that appear in the species occurrence records available from the resource. The data elements table lists those fields and some metadata about them.

You should review the data elements table:

Contents Preview Metadata								
Name	XSDType	ArcGISType	Searchable	Returnable	Namespace	Location	Descr	
RecordURL	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	Gives the web address of the page where	
Source	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	Indicates who gave the record to the data p	
Citation	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	Indicates how this record should be attrib	
Subgenus	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	The subgenus name of the organism	
StartYearCollected	xsd:gYear	SHORT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
EndYearCollected	xsd:gYear	SHORT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
StartMonthCollected	xsd:nonNegativeInteger	LONG	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
EndMonthCollected	xsd:nonNegativeInteger	LONG	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
StartDayCollected	xsd:nonNegativeInteger	LONG	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
EndDayCollected	xsd:nonNegativeInteger	LONG	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
StartJulianDay	xsd:nonNegativeInteger	LONG	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
EndJulianDay	xsd:nonNegativeInteger	LONG	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
StartTimeOfDay	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
EndTimeOfDay	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events tha	
TimeZone	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	Indicates the time zone for the Time of Day	
StartLongitude	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events bet	
EndLongitude	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events bet	
StartLatitude	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events bet	
EndLatitude	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For samples/observations/record events bet	
Start_EndCoordinatePrecision	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	An estimate of how tightly the locality was s	
DepthRange	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	For data sets that have the depth range exp	
Temperature	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	The temperature recorded with the collectio	
LifeStage	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	Indicates the life stage present. Will require	
ObservedIndividualCount	xsd:nonNegativeInteger	LONG	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	The number of individuals (abundance) found	
SampleSize	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	Sample_size: the size of the sample from w	
ObservedWeight	xsd:decimal	DOUBLE	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	The total biomass found in a collection/recor	
GMLFeature	xsd:string	TEXT	1	1	http://www.iobis.org/obis	http://www.iobis.org/obis/obis.xsd	Geographic Markup Language(GML) descrip	
DateLastModified	xsd:dateTime	DATE	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	ISO 8601 compliant stamp indicating the date	
InstitutionCode	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	A "standard" code identifier that identifies th	
CollectionCode	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	A unique alphanumeric value which identifie	
CatalogNumber	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	A unique alphanumeric value which identifie	
ScientificName	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	The full name of lowest level taxon the Catal	
BasisOfRecord	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	An abbreviation indicating whether the reco	
Kingdom	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	The kingdom to which the organism belongs	
Phylum	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	The phylum (or division) to which the organi	
Class	xsd:string	TEXT	1	1	http://digir.net/schema/conceptual/darwin/2003/1.0	http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd	The class name of the organism	

Each row of this table corresponds to a field of the records available from the server's resources. By default, the **Search DiGIR Records and Create Points** will attach all of these as attributes to the points it creates. In DiGIR terminology, these fields are formally called *data elements*, and their characteristics are defined in *conceptual schemas*. An important part of DiGIR's design is that it does not require the resources on a server to use the same conceptual schemas. An advantage of this design is that it allows data contributors to disagree about what fields should be in their biogeographic databases; they can each define their own fields, if desired. It also allows the databases to evolve over time without requiring DiGIR itself to be changed. A disadvantage is that it pushes complexity onto you, the consumer of the data: if different resources use different conceptual schemas, it is up to you to understand the implications and adjust your analysis accordingly.

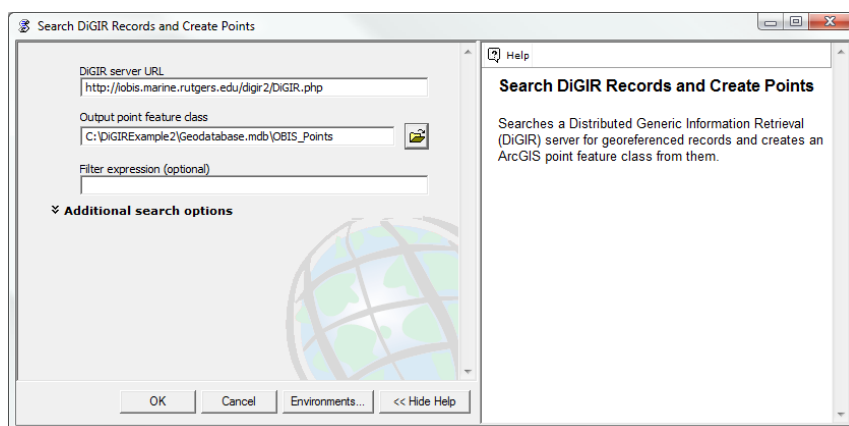
In practice, a relatively small number of conceptual schemas are in widespread use. The most common is called Darwin 2, defined in the document <http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd>. OBIS defines a conceptual schema that extends Darwin 2 by adding 27 additional fields, defined in <http://www.iobis.org/obis/obis.xsd>. The **Big Dig** website lists the conceptual schemas that were in use in February 2008.

The important fields of the data elements table are:

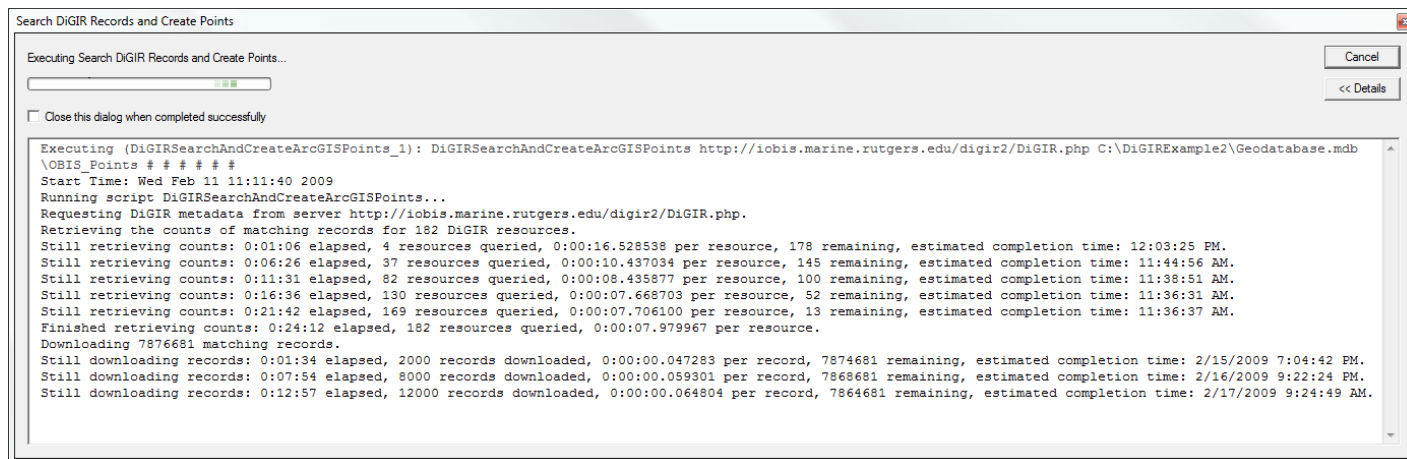
- **Name** specifies the name of the data element. It will be created as an attribute of the points using this name. The points may contain a slightly different name if the data element name is not a legal ArcGIS attribute name. For example, the name Order may be changed to Order_ because Order is a reserved word in database systems. If the points are a shapefile, all of the names will be truncated to 10 characters, the maximum allowed attribute name in a shapefile.
- **ArcGISType** specifies the ArcGIS data type that will be used when the attribute is created for this data element.
- **Searchable** indicates whether you can use the data element name in the **Filter expression** parameter of the **Search DiGIR Records and Create Points** tool. A value of 1 indicates that you can use it.
- **Returnable** indicates whether the data element can be returned by the server. A value of 1 indicates it is returnable. Data elements that are not returnable are intended only for searching. I only know of one such element: the BoundingBox element of the Darwin 2 schema. I was never able to locate the documentation for this element, so I never use it. Instead, I always use the Latitude and Longitude elements.
- **Descr** contains the description of the data element, taken from the conceptual schema that defines it.

Downloading points from a DiGIR server

After you have gained an understanding of the resources available on the DiGIR server, you are ready to download some points using the **Search DiGIR Records and Create Points** tool. This tool only has two required parameters, the DiGIR server URL and the name of the output point feature class to create:



If you just specify these two parameters, the tool will try to download all of the georeferenced records offered by the server. You can monitor the tool's processing in the ArcGIS progress window:



In this example, I attempted to download all of the records from the OBIS server. There are a few things you should note about this example:

- Before the tool can download the records, it has to obtain the count of records that match your request. To do this, it must query the server once for each resource available on the server. In this example, there were 182 resources. Depending on the speed of the server, this can take quite some time. In this example, it took 24 minutes. You can speed this up by restricting your request to specific resources, as described below.
- In this example, the OBIS server reported that 7876681 records are available. This is significantly less than the 16.7 million records that the OBIS home page reports are available. Because the tool creates points, it only downloads records that include geographic coordinates. At the time I placed this request, the OBIS server was undergoing some kind of maintenance and a substantial number of records were missing coordinates.

TO BE CONTINUED...