## Contents

**Getting Started with Scholars@Duke Data** 1
Who is this Guide for? 1
Why consume Scholars@Duke data? 1
How up-to-date is Scholars@Duke data? 1
What limitations may I find in Scholars@Duke data? 1

**Using Scholars@Duke data** 2
Widgets 2
SPARQL Endpoint 5
Drupal Module 6
Sites@Duke (WordPress) 6

**Elements Reporting** 7
Features of the Elements Reporting Tool 7
Getting Started with the Elements Reporting Tool 7
Advanced Use of Elements Reporting 8
Getting Help and More Information 8

**Widgets Data Dictionary (Complete feed, JSON)** 9
Getting Started with Scholars@Duke Data

Who is this Guide for?
Beyond the export of data from a single profile, Scholars@Duke provides several methods for exporting data across groups of individuals and entire organizations. This guide is written for anyone who would like to aggregate and/or export larger data sets from Scholars@Duke.

Why consume Scholars@Duke data?
Scholars@Duke displays researcher data sourced from official systems of record at Duke University. From each of these source systems, researchers and their delegates have the ability to maintain and customize the information that they prefer to display on their public profile. There are several benefits for re-using the information that is provided in Scholars@Duke:

- One-stop shop for Duke faculty data
- Authoritative information about Duke researchers
- Data is consistently structured across all researchers in all organizations
- The underlying linked data model facilitates data aggregation and complex network analyses
- Customizable and dynamic data feeds available in multiple formats
- Minimizes data maintenance in downstream systems and applications

How up-to-date is Scholars@Duke data?
Scholars@Duke is refreshed nightly from its source systems. Researchers and their delegates can trigger real-time data refreshes via the “Manage My Profile” feature. When a new faculty appointment is created or ended, the updated information will be indicated in Scholars@Duke in 24 hours.

What limitations may I find in Scholars@Duke data?
The data available in Scholars@Duke is ideal for most data consumer needs including downstream web applications and analyses. However, for tasks such as CV generation or non-public reporting, Scholars@Duke will provide a starting point from which additional information may need to be added.

- Hidden/Private Data Not Available
  Scholars@Duke only displays data that is approved to be publically accessible. For the most part, researchers and their delegates are able to designate which elements of their profile they want hidden. This could include individual data elements, such as a grant or course, or entire sections such as Publications. Grants from private sponsors are hidden by default pending review of publicity restrictions.

- Some Legacy Data Not Available
  Some data sections in Scholars@Duke such as Appointments & Affiliations, and Courses provide only the most recent data. Expired appointments and courses older than 3 years old are not be available.

- Subject Heading (previously titled ‘Keywords/Research Interests’) are limited to selected vocabularies
  In support of the linked data structure underlying Scholars@Duke, official controlled vocabularies are used whenever possible to facilitate data aggregation and data sharing. Keywords are populated on a profile both manually by the researcher and automatically through the publisher keywords tagged on each publication. In both cases (manual and auto-generated), keywords are selected from preferred terms in the MeSH and Library of Congress vocabularies. Due to limitations in these vocabularies for certain disciplines, the Keywords data may be less informative/accurate for some profiles than others. Additional, more-specific, keywords for a researcher can sometimes be found in their Overview.
Current faculty and select researchers available

Scholars@Duke provides web profiles for all current Duke regular-rank and non-regular rank faculty members as well as some non-faculty researchers including Duke students and staff.

Scholars@Duke profiles are automatically generated and removed for regular and non-regular rank faculty members. Faculty must maintain a current faculty appointment at Duke in order for their profile to remain active in the Scholars@Duke system. Non-faculty researcher profiles are created manually on an opt-in basis. Non-faculty profile must be added by a Scholars@Duke organizational editor based on the discretion of that organizational unit.

Using Scholars@Duke data

Depending on your data needs, there are three methods of aggregating and consuming data from Scholars@Duke.

- Widgets
- SPARQL Endpoint
- Drupal module

The following section compares & contrasts each method, along with the reporting functions in Elements, the online service provided by Duke Libraries for managing publications.

Widgets

Widgets Overview

Widgets are a low-maintenance solution for consuming data about an entire organization. Widgets are accessible from a person’s or organization’s profile page by clicking the “Add Data to my Website” button. While the organizational widgets provide a list of personnel affiliated with the organization, widgets will also work with a customized list of individuals.

- Embed code available for an individual includes publications, artistic works, courses, grants, appointments, addresses, overview, webpages, keywords, geographical foci, photo, and contact information. Additional data elements for an individual are available as JSON using the ‘Complete feed’ option.
- Embed code available for an organization includes people and grants. Photos and overviews are also available as JSON using the ‘Advanced’ option.
- Widgets include additional data attributes that may not be apparent from a person’s profile page (ex. appointment type, organizational unit numbers, and favorite publications).
- Data can be customized to return a subset of results (by count or within a date range).
- Dynamic data feed. Embed code available so that data will auto-refresh nightly.
- Available data formats include JSON, JSONP, HTML, and JavaScript.
Getting Started with Widgets

Anyone with a Duke NetID can access the widgets by clicking the “Add Data to My Website” button on any Scholars profile or organization page. More information on the using Widgets Builder is available in the Scholars@Duke User’s Guide.

Advanced Use of Widgets

One way to pull data for an entire organization is to write a script that loops over the individual person widgets using the list of people obtained in an organizations people list. This method works with a list of unique URIs, Duke Unique IDs, or NetIDs. These identifiers can be substituted into the following URL in order to access the JSON feed for an individual:

https://scholars.duke.edu/widgets/api/v0.9/people/complete/all.json?uri=https://scholars.duke.edu/individual/per9294252

 Scholars@Duke Person URI – The person identifier (text in green) is specific to Scholars@Duke. This identifier is what is listed in the organizational widgets Person List and is also a part of the profile page URL.
 Net ID – For non-faculty researchers, the Scholars@Duke Person URI is the text “per” followed by their Net ID rather than a 7 digit numerical value.
 Duke Unique ID – rather than using the person identifier, the widgets URL will also resolve using the Duke Unique ID (for faculty members only).

Publications Metadata in the Widgets

Widgets provides a wealth of parsed publications metadata for downstream applications. Some examples of this metadata include “abstract”, “PubMed ID”, “digital object identifier”, and “full text link”. All of the metadata required to build a standard publication citation can be accessed as individual data elements via the widgets. Alternatively, the widgets provide 4 pre-formatted citation (APA, MLA, Chicago, and ICMJE) for each publication. These pre-formatted citations include a link to the publication’s page on the Scholars@Duke site. An example of the pre-formatted citations can be seen below:
Note: Currently, the citation structure in Elements does not allow for the full first name of an author. This causes some limitations when trying to format for certain citation styles.

In Profile Manager (click “Manage My Profile”), anyone with a Scholars@Duke profile has the ability to set one of these four citation options to be their preferred citation style. Setting this preference will determine how citations are displayed on a person’s profile page and which style is used in the widgets HTML embed code. For those that have not selected a preferred citation style, the default style is the Chicago style. Whether or not someone has selected a preferred citation style, the widgets JSON feed will include all four citation styles. Therefore, when using the JSON data, it will be up to the data consumer to note which style (if any) has been selected as the preference. The preferred citation style will appear in the JSON feed as seen below. If no preference has been set, the preferredCitationFormat attribute will not exist.

```
"preferredCitationFormat": "http://v4o.duke.edu/v4o/ontology/duke-extension#chicagoCitation"
```
SPARQL Endpoint

SPARQL Endpoint Overview

The SPARQL endpoint is a query service that implements the SPARQL Protocol for RDF and provides SPARQL query processing for RDF data available on the open internet. The Scholars@Duke SPARQL endpoint utilizes the full potential of querying RDF (linked) data by providing innumerable possibilities for aggregating and querying the data. This option may be considered if data needs are required beyond what is provided by the widgets.

Features of the SPARQL endpoint include:
- Similar to SQL queries, SPARQL queries can answer more targeted questions about the data in Scholars@Duke.
- Queried yield static results. Unlike embedding widgets code, the query will need to be run periodically to refresh the data.
- Formats include JSON, XML, Text, CSV, and TSV.

Getting Started with the SPARQL Endpoint

Contact the Scholars@Duke team (scholars@duke.edu) to inquire about use of the SPARQL endpoint. Someone will be in contact with you to assist in evaluating your data needs and writing the appropriate query. Using the SPARQL endpoint requires a NetID and an API key.

Advanced Use of the SPARQL Endpoint

SPARQL queries are specific to our underlying ontology. Details about the ontology are available at https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology

Available ontological attributes can be browsed using the SPARQL inspector: https://sparql.scholars.duke.edu/inspector-request.html

The Scholars@Duke team is available to assist in crafting and optimizing queries. Some basic resources for learning SPARQL can be found here:
**Drupal Module**

As a way to support Drupal-based development at Duke University, Duke University Web Services (DWS) has developed a custom Drupal module that makes Scholars@Duke data available for Drupal using the widgets. For more information about the code and available support, please visit http://webservices.duke.edu/drupalcode/.

**Sites@Duke (WordPress)**

Widgets data is now available for use in a Sites@Duke WordPress page. From a Scholars@Duke profile page:

- Go to ‘Add Data to my Website’
- Select the dataset you would like to add to WordPress
- Click Advanced
- Click Javascript format. This will open up a new window with the JavaScript code.
- Copy the URL of this Javascript page

On your sites.duke.edu site, go to the post or page where you want the Scholars info to appear and paste. It does not matter if you’re on the Visual or Text tab. You should see the info from Scholars@Duke instantly appear. Save/publish like usual in sites.duke.edu.
Elements Reporting

Symplectic Elements is the publication management system at Duke University and the source system for all publications in Scholars@Duke. This application is managed by Duke Libraries and has reporting capabilities for data consumers who are interested in publications data for one or more organizations. For help with accessing reporting tools in Elements, please visit http://scholarworks.duke.edu/elements.

Features of the Elements Reporting Tool

The Elements reporting tool enables aggregation and filtering for customizing queries. Data exports are available in CSV or Reference Manager/Endnote formats.

Getting Started with the Elements Reporting Tool

Elements requires a Duke NetID. To use the reporting tool, you’ll need additional “statistician” (reporting) rights that must be granted by Duke Libraries. Please contact elements@duke.edu for more information and to request this access.
Advanced Use of Elements Reporting

Publications metadata available:

<table>
<thead>
<tr>
<th>ID</th>
<th>Open access</th>
<th>PDS ID</th>
<th>Proprietary ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>ISSN</td>
<td>Translator</td>
<td>Source</td>
</tr>
<tr>
<td>Username</td>
<td>Issue</td>
<td>Citation count (Scopus)</td>
<td>Other sources present</td>
</tr>
<tr>
<td>Email</td>
<td>Journal Of Published proceedings</td>
<td>Citation count (Web of Science (LItE))</td>
<td>arXiv record exists?</td>
</tr>
<tr>
<td>User's Proprietary ID</td>
<td>Keywords</td>
<td>Citation count (Europe PubMed Central)</td>
<td>British Library record exists?</td>
</tr>
<tr>
<td>Primary group descriptor</td>
<td>Language</td>
<td>SNIP Journal name</td>
<td>DIMI RN record exists?</td>
</tr>
<tr>
<td>Primary group</td>
<td>Conference place Of Country</td>
<td>SNP rank</td>
<td>DIMI JP record exists?</td>
</tr>
<tr>
<td>Is current staff</td>
<td>Medium</td>
<td>ERI journal name</td>
<td>CrossRef record exists?</td>
</tr>
<tr>
<td>Visible</td>
<td>Name of conference</td>
<td>DBLP rank</td>
<td>Digital share exists?</td>
</tr>
<tr>
<td>Favourite</td>
<td>Notes</td>
<td>SJR journal name</td>
<td>Europe PubMed Central record exists?</td>
</tr>
<tr>
<td>Publication type</td>
<td>Article number</td>
<td>SIGIR rank</td>
<td>Google Scholar record exists?</td>
</tr>
<tr>
<td>Reporting date 1</td>
<td>Pagination (start page)</td>
<td>Indexed in DOAJ</td>
<td>Google Books record exists?</td>
</tr>
<tr>
<td>Reporting date 2</td>
<td>Pagination (end page)</td>
<td>DOAJ CC Licence</td>
<td>Google Scholar record exists?</td>
</tr>
<tr>
<td>Abstract</td>
<td>Pagination (page count)</td>
<td>Romeo-colour</td>
<td>Google Scholar record exists?</td>
</tr>
<tr>
<td>Addresses</td>
<td>PI</td>
<td>Romeo-colour last updated when</td>
<td>Local Source 1 record exists?</td>
</tr>
<tr>
<td>Authors</td>
<td>Place of publication</td>
<td>Files in Preproduction DukSpace instance</td>
<td>Manual record exists?</td>
</tr>
<tr>
<td>Link to item</td>
<td>Publication date</td>
<td>Preproduction DukSpace instance public URL</td>
<td>Mendeley record exists?</td>
</tr>
<tr>
<td>DOI</td>
<td>Status</td>
<td>Licence granted to Preproduction DukSpace instance</td>
<td>ORCID record exists?</td>
</tr>
<tr>
<td>Edition</td>
<td>Publisher</td>
<td>Item status in Preproduction DukSpace instance</td>
<td>PubMed record exists?</td>
</tr>
<tr>
<td>Editors</td>
<td>Publisher URL</td>
<td>Embargo period requested in Preproduction DukSpace instance</td>
<td>RePEc record exists?</td>
</tr>
<tr>
<td>eISSN</td>
<td>Conference start date</td>
<td>Files in Production DukSpace instance</td>
<td>Scival Export record exists?</td>
</tr>
<tr>
<td>External identifiers</td>
<td>Title</td>
<td>Production DukSpace instance public URL</td>
<td>Scopus record exists?</td>
</tr>
<tr>
<td>Conference finish date</td>
<td>Sub types</td>
<td>Licence granted to Production DukSpace instance</td>
<td>Web of Science (LItE) record exists?</td>
</tr>
<tr>
<td>ISBN-10</td>
<td>Volume</td>
<td>Item status in Production DukSpace instance</td>
<td>Web of Science (LItE) record exists?</td>
</tr>
<tr>
<td>ISBN-13</td>
<td>Type of conference publication</td>
<td>Embargo period requested in Production DukSpace instance</td>
<td>Canonical journal title</td>
</tr>
</tbody>
</table>

Getting Help and More Information

To submit comments, questions, bugs, or enhancement requests concerning Scholars@Duke, please submit a help ticket (https://about.scholars.duke.edu/content/submit-scholarsduke-help-ticket). To be added to the Scholars@Duke data consumers group listserv, please send an email to damaris.murry@duke.edu and put “DCG listserv addition” in the subject line.

For specific information about support policies for Duke website administrators as well as developers consuming scholars data, refer to the document, “Support for Data Consumers” on the Scholars Support page (http://about.scholars.duke.edu/support-duke-faculty-delegates).
Widgets Data Dictionary (Complete feed, JSON)

November 2015

{

uri: the unique identifier for a person in Scholars@Duke. Faculty URIs include a value that is their hashed Duke Unique ID. Non-faculty URIs include their Duke Net IDs. The URI redirects to the URL of the Scholars@Duke profile.

vivoType: the type of affiliation with Duke. This is what determines the structure of the URI.

label: the Professional Name (Last, First, Middle) designated in dFac

title: the Preferred (aka “highest precedence”) title as designated in dFac. This is the title in which the attribute of “rank” has the lowest value.

publications each publication feeding in from Elements (elements.duke.edu) will have some of the following metadata. Metadata with an asterisk will be present for each publication, while the others are optional.

{

uri: the unique identifier for a publication in Scholars@Duke. The six-digit value after “pub” represents the identifier that comes from Elements. The URI redirects to the URL for the publication’s entity page in Scholars@Duke.

vivoType: the type of publication as designated in Elements. The list of publication types is fairly static, however it is possible for a new type to be added to Elements. The types currently available are:

bibo:AcademicArticle                        duke:JournalIssue
bibo:Book                                    duke:OtherArticle
bibo:BookSection                             vivo:ConferencePaper
bibo:EditedBook                              vivo:Dataset
bibo:Report                                  vivo:NewsRelease
bibo:Thesis                                  vivo:Review
duke:BookSeries                              vivo:Software
duke:DigitalPublication

label: the title of the publication

attributes: {this section contains 4 pre-formatted publication citations, as well as the individual data elements that make up these citations

  icmjeCitation: International Committee of Medical Journal Editors citation style
  mlaCitation: Modern Language Association citation style
  isFavorite: this attribute is set to “true” when an author has marked their publication as a “favorite” in Elements.

authorList: list of all authors as “Last Name, First Initial”
datetime: publication date URI as ‘YYYYMMDD’

pmid: PubMed ID

apaCitation: American Psychological Association citation style

year: publication date as “YYYY-MM-DDT00:00:00”

endPage: end page of publication (typically for an article or book section).
publicationSource: the source of the publication record. Metadata will vary slightly between the various sources. A preferred source (per publication) can be selected in Elements. The default preferred source is PubMed / European PubMed. Possible sources are
"arxiv" (ArXiv)
"cinii-english" (Scholarly and Academic Information Navigator)
"cinii-japanese"
"c-inst-1" (Fuqua School of Law publication repository)
"crossref" (CrossRef)
"dblp" (Computer Science Bibliography)
"epmc" (European PubMed)
"google-books" (Google Books)
"manual" (manually added)
"pubmed" (PubMed)
"repec" (Research Papers in Economics)
"scival" (SeiVal/REACH NC)
"scopus" (Scopus)
"ssrn" (Social Science Research Network)
"wos-lite" (Web of Science)

chicagoCitation: Chicago citation style
authorshipType: relationship between person and publication. Values are “Authorship”, “Editorship”, or “Translatorship”.
issue: journal issue
subtypes: while the vivoType represents the format of the publication, the subtype relays the genre and/or content of the publication.
publicationVenue: The journal represented as unlinked text (when the ISSN is not identified) or a URI (when the ISSN is identified)
publishedIn: publisher name
startPage: start page of publication (typically for an article or book section).
do: digital object identifier
volume: journal volume

awards: awards & honors that are manually entered in Scholars@Duke. Because the data is manually entered, this section may also include grants that were not submitted to SPS (Duke’s official grants database).

{ uri: The unique identifier for an award in Scholars@Duke.}
vivoType: an award as defined in the VIVO-ISF ontology (vivo:Award)
label: a concatenation of the award name & awarding body
attributes:
  - serviceType: Represents the level of award. Values can be “Department”, “International”, “National”, “Other”, “School”, “State”, or “University”.
  - name: The name of the award
  - datePrecision: the Scholars URI for the date
  - awardedBy: the name of the awarding body
  - awardedByUri: the Scholars URI for the awarding body
  - date: the date in text form
}

artisticWorks: [non-print outputs that represent a faculty member’s primary scholarship. These are works that are parallel to publications and not adequately captured in the publication citation style]
grants: [select grants from Duke’s Sponsored Project System (SPS). If applicable, one grant will represent multiple renewal years, as well as any mini grants that associated with it. Grants with publicity restrictions are not included.
  - uri: the unique identifier for a grant in Scholars@Duke. The value after the “gra” represents the proposal ID. In the case of a renewal grant, the URI value will automatically update to the most current proposal ID.
    - vivoType: categories of grants. These include
      Clinical Trial Grant
      Conference Grant
      Construction Renovation Grant
      Equipment Grant
      Fellowship Grant
      Institutional Support Grant
      Institutional Training Grant
      Public Service Grant
      Research Grant
    - label: name of grant
    - attributes:
      - endDate: project period end date for the proposal (most recent proposal if grant has been renewed)
      - roleName: person’s role on the grant
      - awardedBy: name of the grant sponsor
      - administeredBy: name of the Duke department that received the grant
      - awardedByUri: URI of the grant sponsor
      - administeredByUri: URI of the Duke department that received the grant
      - startDate: project period start date for the proposal (first proposal if grant has been renewed)
courses: [ courses taught in the past 3 years

{ uri: the unique identifier for the course. The value after “course” represents the course number. The URI for renumbered courses will automatically be updated for all course years.
  vivoType: vivo:Course
  label: a concatenation of the course name and course title.
  attributes: {
    role: the URI representing the role of a particular instructor on a course
    roleName: currently, this field is a duplication to the course label (concatenation of the course name and course title)
  }
}

professionalActivities: []

positions: [ faculty appointments and center memberships contained in dFac. This section also includes non-faculty member’s HR title and non-faculty affiliations as designated in Scholars@Duke.

{ uri: unique identifier for a position
  vivoType: the type of affiliation a position represents. This could be
  Primary Position
  Secondary Position
  Joint Position
  Faculty Administrative Position
  Professorship
  Membership Position
  Student Position
  Non-Faculty Academic Position
  Affiliate Position
  Faculty Position
}
**label**: appointment title or non-faculty HR title. For students this will always be “Student”

**attributes**: {

  **startDatetimeUri**: appointment start date URI
  **schoolLabel**: name of the school for which the appointment exists
  **organizationLabel**: name of the department for which the appointment exists
  **organizationUri**: department URI
  **rank**: display order for titles. These values have a default setting and can be edited in dFac using the Title Precedence form.
  **startYear**: appointment start date in YYYY-MM-DDT:00:00:00 format
  **startYear**: appointment end date in YYYY-MM-DDT:00:00:00 format
  **dateUri**: appointment start and end as a time interval URI
  **personUri**: URI for person who occupies this position
  **endDatetimeURI**: appointment end date URI
  **schoolUri**: URI for the school

}

**addresses**: [ there are 2 addresses in Scholars@Duke--the mailing address and the office location

  
  **uri**: the unique identifier for a person’s address, either work_mailing, or work_location.
  **vivoType**: using the vcard ontology, office location are categorized as vcard:Location, while mailing addresses are vcard:Address
  **label**: concatenation of street, city, state, postal code
  **attributes**: {
    **city**: name of city
    **state**: initials of state
    **postalCode**: postal code
    **personUri**: person associated with the office location/address
    **address1**: line 1 of a street address
    **address2**: line 2 of a street address
  }

}

**educations**: [ a faculty member’s education and professional training as displayed in dFac. This section is only available for faculty who are active in dFac.

  
  **uri**: the unique identifier that specifies the university name, graduation year, and degree earned
  **vivoType**: ontological categorization of education URI (vivo:EducationalProcess)
  **label**: concatenation of degree earned + graduation year
attributes: {
    degreeUri: URI for degree type
    endDate: graduation/training year as “YYYY-DD-MMM00:00:00”
    degree: degree abbreviation
    organizationUri: the unique identifier for the university
    institution: name of the university
    dateTimeUri: graduation/training year as time interval URI
    personUri: the URI for the person associated with the education
    endUri: graduation/training year as end date URI
}

researchAreas: [subject headings from MesH & Library of Congress that describe a person’s research & areas of expertise

    uri: subject heading URI
    vivoType: ontological categorization of subject headings is skos:Concept
    label: name of the subject heading
    attributes: {
        personUri: the person associated with that subject heading
    }

webpages: websites that were manually entered onto a person’s profile (typically includes links to a CV, department site, personal site, or Google Scholar page, etc.)

geographicalFocus:

    uri: the unique identifier for a geographic region
    vivoType: the ontological category for a geographic region
    label: name of the geographic region
    attributes:
        focusTypeLabel: specification of a person’s scholarly relationship to a particular geographic region
        focusType: URI for the relationship type between a person and a geographic region
        personUri: URI for the person

attributes:{
    preferredTitle: the title in the “Positions” section of the widgets that has the lowest “rank” value (aka highest precedence value)
    phoneNumber: phone number from enterprise directory
    primaryEmail: email from enterprise directory
    middleName: middle name (if available)
    imageUri: URI for page that links to full image and thumbnail downloads

Scholars@Duke Data Consumer’s Guide
mentorshipAvailabilities: paragraph about availability to mentor
netid: net ID
alternateId: Duke Unique ID
prefixName: title prefix
imageDownload: link to full profile image
lastName: last name
firstName: first name
imageThumnailUri: URI for page that links to thumbnail download
imageThumbnailDownload: link to thumbnail of profile image
overview: the overview paragraph that summarizes a person’s research and research interests.