

Table of Contents

1. [Overview](#)
2. [XML Schema Breakdown](#)
3. [NSF Organization Hierarchy](#)
4. [Terminology](#)
 1. [American Recovery and Reinvestment Act \(ARRA\) of 2009](#)
 2. [Funding Opportunity Announcement](#)
 3. [Assistance Awards](#)
 1. [Grant](#)
 1. [Standard Grant](#)
 2. [Continuing Grant](#)
 2. [Cooperative Agreement](#)
 3. [Cost Reimbursement Award](#)
 4. [Fixed Amount Award](#)

Overview

The raw award data is warehoused by the NSF and obtainable through the nsf.gov website, [here](#). There is also a RESTful endpoint which can be used to download the data by year:

```
http://www.nsf.gov/awardsearch/download?DownloadFileName=`2014`&All=true
```

Each file that is downloaded is a zipped XML file with a listing of all awards for that year. So for instance, a POST request to the above URI will return a zipped XML file that contains all the award data for the year 2014. The schema for the XML data can be seen [here](#). It will be elaborated upon below in order to formulate ideas for and detail assumptions made when parsing the awards into a cleaned form which will be stored in a SQL database. The goal of this effort is to produce a high quality representation of the data with minimal redundancy and maximum clarity, while ensuring the XML representation of the awards is interpreted correctly.

XML Schema Breakdown

Each year is downloaded as a zip file and is named using the year: `<year>.zip`. Each zip file contains a bunch of XML files, each containing info for exactly one award. These files are named using the award ID: `<awardID>.xml`. Each XML file has a `rootTag`, followed by an `<Award>` tag which contains the following elements:

1. **AwardID** (*int*): The unique ID of the award.
2. **AwardTitle** (*string*): The title of the award.
3. **AbstractNarration** (*string*): An overview of what the research is seeking to do.
4. **AwardEffectiveDate** (*dateTime*): The date the award funding starts.
5. **AwardExpirationDate** (*dateTime*): The date the award funding ends.
6. **MinAmdLetterDate** (*dateTime*): The first date the award was amended.
7. **MaxAmdLetterDate** (*dateTime*): The last date the award was amended.
8. **AwardAmount** (*int*): The amount of money awarded to date.
9. **ARRAAmount** (*string*): Portion of AwardAmount funded by the American Recovery and

Reinvestment Act (*ARRA*).

10. **AwardInstrument** (*sequence*): Listing of classifications for this award.
 1. **Value** (*string*): A particular classification (e.g. "Standard Grant", "Cooperative Agreement", "Contract").
11. **Organization** (*sequence*): The NSF organization(s) funding the grant.
 1. **Code** (*int*): No one seems to know what this actually represents. Perhaps it can be a challenge for some daring historian/researcher to discover.
 2. **Directorate** (*sequence*): Listing of directorates funding this award.
 1. **LongName** (*string*): Name of directorate.
 3. **Division** (*sequence*): Listing of divisions funding this award.
 1. **LongName** (*string*): Name of division.
12. **ProgramElement** (*sequence*): Listing of programs funding this award.
 1. **Code** (*int*): Unique ID of the program funding.
 2. **Text** (*string*): Name of the program funding.
13. **ProgramReference** (*sequence*): Listing of programs intellectually related to the subject matter of this award.
 1. **Code** (*int*): Unique ID of the program referenced.
 2. **Text** (*string*): Name of the program referenced.
14. **ProgramOfficer** (*sequence*): A listing of all Program Officers responsible for this award.
 1. **SignBlockName** (*string*): The name of the Program Officer.
15. **Investigator** (*sequence*): A listing of all investigators who have worked on or are working on this award.
 1. **FirstName** (*string*): The first name of the investigator.
 2. **LastName** (*string*): The last name of the investigator.
 3. **EmailAddress** (*string*): The email address of the investigator (optional).
 4. **StartDate** (*dateTime*): The date the investigator started working on this award.
 5. **EndDate** (*dateTime*): The date the investigator stopped working on this award.
 6. **RoleCode** (*int*): The role of the investigator, identified by an integer code. Either "Principal Investigator" or "Co-Principal Investigator".
16. **Institution** (*sequence*): The institution sponsoring this award (PO/Investigator affiliation).
 1. **Name** (*string*): Name of the institution.
 2. **PhoneNumber** (*decimal*): Phone number of the institution.
 3. **CityName** (*string*): Name of the city where the institution is located.
 4. **StreetAddress** (*string*): Name of the street on which the institution is located.
 5. **StateCode** (*string*): The
 6. **StateName** (*string*): Name of the state in which the institution is located.
 7. **ZipCode** (*int*): Zip code of the institution's postal address.
 8. **CountryName** (*string*): Name of the country in which the institution is located.
17. **FoalInformation** (*sequence*): Funding Opportunity Announcement (FOA) reference (to Grants.gov/nsf.gov FOA listing).
 1. **Code** (*int*): Unique ID of FOA.
 2. **Name** (*string*): Name of FOA.

NSF Organization Hierarchy

The NSF is organized into a hierarchy of sub-organizations. There are two top-level types of sub-organizations.

1. Directorate: in charge of multiple divisions.
 1. Division: in charge of multiple programs.

1. Program: each program has only one controlling division.
2. Office: can be organized into multiple subordinate offices and divisions, as well as multiple subordinate programs.
 - Each office/sub-office may have arbitrarily many sub-offices/divisions.
 - Programs encompassed by offices are uniquely controlled by only one office.
 - Offices can be thought of like directories in a file system, where programs are files and sub-offices are sub-directories.
 - Some offices are further divided into branches, but these are not tied to programs, so they are not particularly interesting for a program-centric dataset.

Unique identifiers:

1. Directorate/Division: both uniquely identified by a string abbreviation; see the [NSF orglist](#)
2. Program: uniquely identified by the program code, which is 4 characters, where characters can be digits or letters.
 - programs were not assigned codes before **1975**, so data before then will be difficult to use
 - note that the same program may be identified by a variety of different codes
 - each uniquely identifies the program
 - different codes are generated by the same program if:
 1. it moves from one division/office to another
 2. it's purpose changes significantly
 3. other miscellaneous reasons motivated by accounting needs cause it to
3. Institution: no fundamentally unique identifier; the address/name combo is probably the least ambiguous.

Terminology

Sources: [NSF Proposal and Award Policies and Procedure Guide]

(http://www.nsf.gov/pubs/policydocs/pappguide/nsf08_1/index.jsp#C) | [Grants.gov Glossary](#)

American Recovery and Reinvestment Act (ARRA) of 2009

The economic stimulus package of \$787 billion (Also known as the "Recovery Act", was signed into law by the President on February 17, 2009; it is the economic stimulus package of \$787 billion.

"Making supplemental appropriations for job preservation and creation, infrastructure investment, energy efficiency and science, assistance to the unemployed, and State and local fiscal stabilization, for the fiscal year ending September 30, 2009, and for other purposes".

Funding Opportunity Announcement

A publicly available document by which a federal agency makes known its intentions to award discretionary grants or cooperative agreements, usually as a result of competition for funds.

Funding opportunity announcements may be known as program announcements, notices of funding availability, solicitations, or other names depending on the agency and type of program.

Funding opportunity announcements can be found at [Grants.gov/FIND](#) and on the Internet at the funding agency's or program's website.

Assistance Awards

Awards that entail the transfer of money, property, services or other things of value from the Federal government to a recipient to accomplish a public purpose of support or stimulation. In the case of NSF, assistance awards involve the support or stimulation of scientific and engineering research, science and engineering education or other related activities. NSF is authorized to use grants or cooperative agreements for this purpose.

Grant

A type of assistance award and a legal instrument which permits an executive agency of the Federal government to transfer money, property, services or other things of value to a grantee when no substantial involvement is anticipated between the agency and the recipient during the performance of the contemplated activity. Grants are the primary mechanism of NSF support. NSF awards the following two types of grants.

Standard Grant

A type of grant in which NSF agrees to provide a specific level of support for a specified period of time with no statement of NSF intent to provide additional future support without submission of another proposal.

Continuing Grant

A type of grant in which NSF agrees to provide a specific level of support for an initial specified period of time, usually a year, with a statement of intent to provide additional support of the project for additional periods, provided funds are available and the results achieved warrant further support.

Cooperative Agreement

A type of assistance award which should be used when substantial agency involvement is anticipated during the project performance period. Substantial agency involvement may be necessary when an activity is technically and/or managerially complex and requires extensive or close coordination between NSF and the awardee. Examples of projects which might be suitable for cooperative agreements if there will be substantial agency involvement are: research centers, policy studies, large curriculum projects, multi-user facilities, projects which involve complex subcontracting, construction or operations of major in-house university facilities and major instrumentation development.

Cost Reimbursement Award

A type of grant under which NSF agrees to reimburse the grantee for work performed and/or costs incurred by the grantee up to the total amount specified in the grant. Such costs must be allowable in accordance with the applicable cost principles (e.g., OMB Circular A-21, Cost Principles for Educational Institutions or OMB Circular A-122, Cost Principles for Non-Profit Organizations). Accountability is based primarily on technical progress, financial accounting and fiscal reporting.

Except under certain programs and under special circumstances, NSF grants and cooperative agreements are normally cost reimbursement type awards.

Fixed Amount Award

A type of grant used in certain programs and situations under which NSF agrees to provide a specific level of support without regard to actual costs incurred under the project. The award amount is negotiated using the applicable cost principles or other pricing information as a guide. This type of grant reduces some of the administrative burden and record-keeping requirements for both the grantee and NSF. Except under unusual circumstances, such as termination, there is no governmental review of the actual costs subsequently incurred by the grantee in performance of the project. There typically is a requirement for the grantee to certify that the approximate number of person-months or other activity called for in the grant was performed. Payments are based on meeting specific requirements of the grant and accountability is based primarily on technical performance and results.

Grantee

The organization or other entity that receives a grant and assumes legal and financial responsibility and accountability both for the awarded funds and for the performance of the grant-supported activity. NSF grants are normally made to organizations rather than to individual Principal Investigator/Project Director(s). Categories of eligible proposers may be found in [GPG Chapter I](#).

Principal Investigator/Project Director (PI/PD)

The individual designated by the grantee, and approved by NSF, who will be responsible for the scientific or technical direction of the project. If more than one, the first one listed will have primary responsibility for the project and the submission of reports. All others listed are considered co-PI/PD, and share in the responsibility of the scientific or technical direction of the project. The term "Principal Investigator" generally is used in research projects, while the term "Project Director" generally is used in science and engineering education and other projects. For purposes of this Guide, PI/co-PI is interchangeable with PD/co-PD.

Grants.gov

A storefront web portal for use in electronic collection of data (forms and reports) for federal grant-making agencies through the Grants.gov site.

Acronyms

Co-PI: Co-Principal Investigator

PI: Principal Investigator

PD: Program Director

Co-PD: Co-Program Director

PO: Program Officer