**The New Database
Design for DIA2**
The DIA2 team

The New Database Design for DIA2

Delivered: Friday, June 1, 2012

Database Design

The New Database Design for DIA2

Main Modules in the Database

PK - Primary key; **FK** - Foreign key

1. Hierarchical categorization of documents

▸ **NSFDirectorate** is a lookup table for NSF directorates.

Column	Data type	Description
ID (PK)	INT	A unique identifier for an NSF directorate

Column	Data type	Description
name	VARCHAR(200)	The directorate name in full
abbreviation	VARCHAR(15)	The abbreviation for the given directorate
description	VARCHAR(300)	Additional description about the directorate

‣**NSFDivision** is a lookup table for NSF divisions.

Column	Data type	Description
ID (PK)	INT	A unique identifier for an NSF division
name	VARCHAR(200)	The division name in full
abbreviation	VARCHAR(15)	The abbreviation for the given division

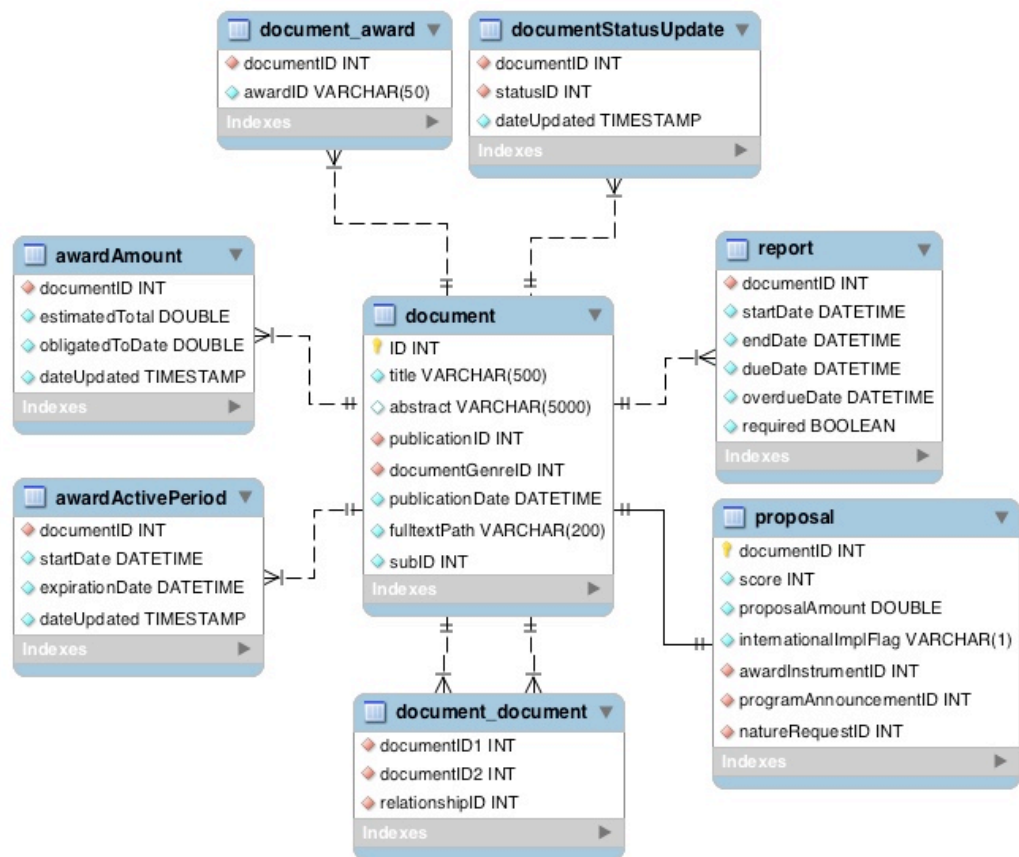‣**NSFProgram** is a lookup table for NSF program.

Column	Data type	Description
ID (PK)	INT	A unique identifier for an NSF division
name	VARCHAR(200)	The division name in full
abbreviation	VARCHAR(15)	The abbreviation for the given division
code	VARCHAR(100)	The NSF program element code

‣**document_NSFDirectorate, document_NSFDivision, and document_NSFProgram** are join tables that associate NSF documents to their corresponding directorates, divisions, and programs.

‣**document** is the generic table that contains any kind of document including NSF grant proposals, annual reports, final reports, reviews, and academic publications.

Column	Data type	Description
ID (PK)	INT	A unique identifier for an NSF division
title	VARCHAR(500)	Document title
abstract	VARCHAR(5000)	Document abstract
publicationID (FK)	INT	The publication venue of the document
documentGenreID (FK)	INT	Document genre (e.g., grant proposal and annual report)
publicationDate	DATETIME	Publication date of the document
fulltextPath	VARCHAR(200)	The file path to the full text
subID	INT	A label to allow insertion of highly similar documents

2. NSF awards, proposals, and reports

▸**awardAmount** contains a history of updates on award amount.

Column	Data type	Description
documentID (FK)	INT	The document ID of the awarded proposal
estimatedTotal	DOUBLE	The total estimated amount of an award
obligatedToDate	DOUBLE	The total amount obligated by the updated date
dateUpdated	TIMESTAMP	The date for this creation/updating of award amount

▸**awardActivePeriod** contains information about the active period of an award.

Column	Data type	Description
documentID (FK)	INT	The document ID of the awarded proposal
startDate	DATETIME	The start date of an award
expirationDate	DATETIME	The expiration date of an award
dateUpdated	TIMESTAMP	The date for this creation/updating of award active period

▸**proposal** includes proposals that have been submitted to NSF, which may be at different stages such as submitted, under review, rejected, and awarded. The status update can be found in the documentStatusUpdate table. The amount requested in a proposal may also be different from the actual amount it eventually receives.

Column	Data type	Description
documentID (PK, FK)	INT	The document ID of a proposal
score	INT	The review score of a proposal
proposalAmount	DOUBLE	The requested dollar amount by the PI
internationalImplFlag	VARCHAR(1)	Indicated by submitter on cover page whether the proposal has international activities associated with it. We set a flag whether this has been identified.
awardInstrumentID (FK)	INT	Award instrument ID
programAnnouncementID (FK)	INT	Program announcement under which proposal was submitted.
natureRequestID (FK)	INT	Nature of request: under which conditions the proposal is being submitted such as new, renewal, and other administrative requests

▸**report** refers to annual reports and final reports associated with NSF awards.

Column	Data type	Description
documentID (FK)	INT	The document ID of a report
startDate	DATETIME	Reporting period start date
endDate	DATETIME	Reporting period end date
dueDate	DATETIME	Reporting period due date
overdueDate	DATETIME	Reporting period overdue date
required	BOOLEAN	Indicate whether the report is required, not required, or waived

▸**document_award** maps an awarded proposal to an award number.

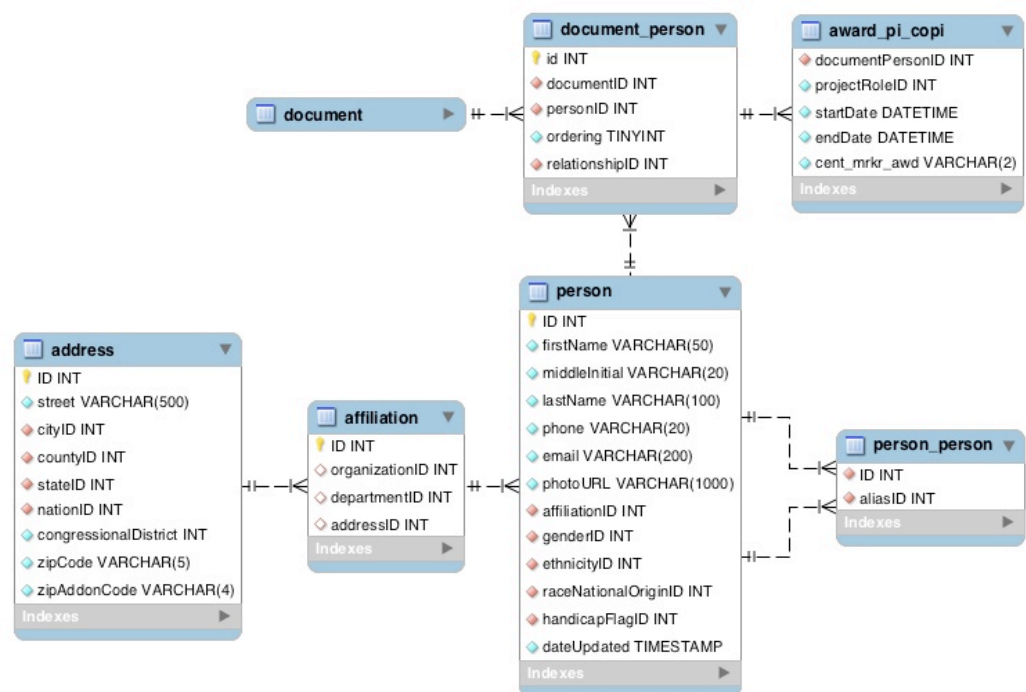| Column	Data type	Description
documentID (FK)	INT	The document ID of the awarded proposal
awardID	VARCHAR(50)	The award ID assigned by NSF

▸**documentStatusUpdate** stores status updates of proposals and annual/final reports. For example, it characterizes cases when a proposal is shifted from "submitted" to "under review". The interval between statuses can be used to compute the dwell stage.

Column	Data type	Description
documentID (FK)	INT	The document ID of the document updated
statusID (FK)	INT	The new status of the document
dateUpdated	TIMESTAMP	The date for this status update

‣**document_document** describes the between-document relationships such as collaborative proposals and proposal-report relationship. It is also used to save the disambiguation results of duplicate documents.

Column	Data type	Description
documentID1 (FK)	INT	The document ID of one awarded proposal
documentID2 (FK)	INT	The document ID of another awarded proposal
relationshipID (FK)	INT	The relationship between document 1 and document 2

3. PI, co-PI, and program officer

‣**document_person** describes the relationship between an document and a person (PI, coPI, program officer, and reviewer).

Column	Data type	Description
id (PK)	INT	The unique ID for the document-person relationship

Column	Data type	Description
documentID (FK)	INT	The document ID
personID (FK)	INT	The person ID associated with the document
ordering	TINYINT	The order of a person: 1 - PI and coPIs are ordered in its original order starting from 2. Other roles such as program officers and reviews have the default value 1.
relationshipID (FK)	INT	Indicate the kind of relationship between the document and the person such as authorship and reviewing.

‣**award_pi_copi** provides additional information about the relationship between PI/coPI and an award.

Column	Data type	Description
documentPersonID (FK)	INT	The ID for the document-person relationship
projectRoleID (FK)	INT	Indicate a person's role on a particular award
startDate	DATETIME	The date when an awardee starts to play the project role
endDate	DATETIME	The date when an awardee discontinues the project role
cent_mrkr_awd	VARCHAR(2)	Unknown

‣**person** contains detailed personal information of an NSF awardee, program officer, and other personnel.

Column	Data type	Description
ID (PK)	INT	The person ID
firstName	VARCHAR(50)	First name
middleInitial	VARCHAR(20)	Middle initial
lastName	VARCHAR(100)	Last name
phone	VARCHAR(20)	Phone number
email	VARCHAR(200)	Email address
photoURL	VARCHAR(1000)	The file path/URL of the headshot image
affiliationID (FK)	INT	Affiliation
genderID (FK)	INT	Gender
ethnicityID (FK)	INT	Ethnicity
raceNationalOriginID (FK)	INT	Race, national origin
handicapFlagID (FK)	INT	Handicap flag
dateUpdated	TIMESTAMP	The creation date of this record

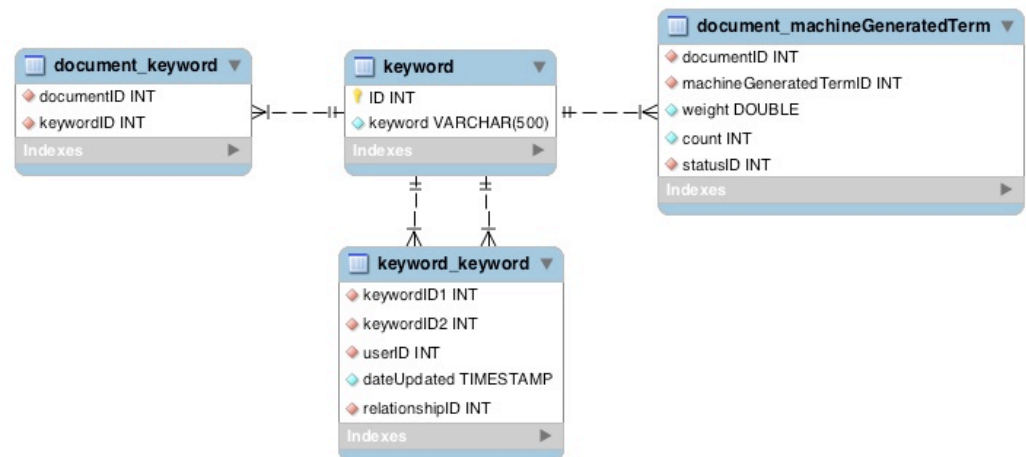‣**person_person** is used to disambiguate duplicate person names.

▸**affiliation** lists the organization, department, and address of a person.

Column	Data type	Description
ID (PK)	INT	Affiliation ID
organizationID (FK)	INT	ID of an organization: university, company, etc.
departmentID (FK)	INT	ID of an department within an organization
addressID (FK)	INT	Address of the given organization and department

▸**address**

Column	Data type	Description
ID (PK)	INT	Address ID
street	INT	Street
cityID (FK)	INT	City
countyID (FK)	INT	County
stateID (FK)	INT	State
nationID (FK)	INT	Nation
congressionalDistrict	INT	Congressional district
zipCode	VARCHAR(5)	Zip code
zipAddonCode	VARCHAR(4)	Zip add-on code

4. Topical description

▸**document_keyword** is a join table of document and author-supplied keyword

▸**document_machineGeneratedTerm** is a join table of document and machine-generated term.

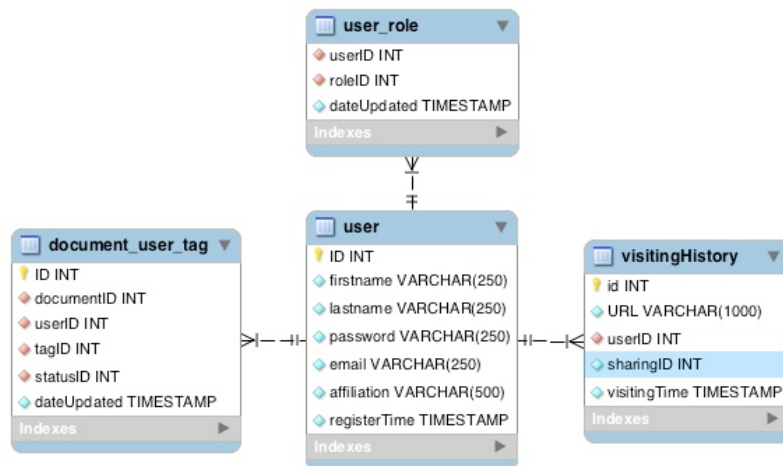| Column	Data type	Description
documentID (FK)	INT	Document ID
machineGeneratedTermID (FK)	INT	Keyphrases generated by algorithms
weight	DOUBLE	A keyphrase's weight computed by algorithms
count	INT	A keyphrase's frequency on a document
statusID (FK)	INT	Indicate whether it is removed

‣**keyword** is a lookup table for both author-supplied keywords and machine-generated terms.

Column	Data type	Description
ID (PK)	INT	Keyword/Keyphrase ID
keyword	VARCHAR(500)	Text

‣**keyword_keyword** describes the between-keyword relationships such as equality and inclusion defined by users.

Column	Data type	Description
keywordID1 (FK)	INT	The first keyword/keyphrase
keywordID2 (FK)	INT	The second keyword/keyphrase
userID (FK)	INT	The user who defines this relationship
dateUpdated	TIMESTAMP	Indicate when the user defines it
relationshipID (FK)	INT	The relationship between the two keywords/keyphrases

5. User

▸**user** saves the information for registered users.

Column	Data type	Description
ID (PK)	INT	User ID
firstname	VARCHAR(250)	First name
lastname	VARCHAR(250)	Last name
password	VARCHAR(250)	Encrypted password for authentication
email	VARCHAR(250)	Email address for login
affiliation	VARCHAR(500)	Affiliation
registerTime	TIMESTAMP	Registration time

▸**user_role** indicates a user's role: normal user, tester, and administrator.

Column	Data type	Description
userID (FK)	INT	User ID
roleID (FK)	INT	Role ID
dateUpdated	TIMESTAMP	Time for granting this role

▸**visitingHistory** tracks users' navigation history and allows sharing of specified visual components.

Column	Data type	Description
id (PK)	INT	ID
URL	VARCHAR(1000)	Page URL visited
creatorID (FK)	INT	ID of the user who visit the page
sharingID (FK)	INT	Sharing options: public, private, and by link
createdTime	TIMESTAMP	Visiting time

▸**document_user_tag** stores users' tags associated with documents.

Column	Data type	Description
ID (PK)	INT	ID
documentID (FK)	INT	The document being tagged
userID (FK)	INT	The user who tags the document
tagID (FK)	INT	Tag content
statusID (FK)	INT	The tag status (regular or removed)

Column	Data type	Description
dateUpdated	TIMESTAMP	Tag creation time

6. ERD overview