

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

Instructions for Collecting Person Information

1 Overview

The issue of Key Personnel, how their information is submitted (or not submitted) via the XML data stream, and how they relate to Commons account information has been an area of intense debate during CGAP Pilot Phase I. So much so, that we have decided to revise the manner in which information on Key Personnel (including the Principal Investigator) shall be collected during CGAP Pilot Phase II (and beyond).

2 Principal Investigator

There are certain issues that are specific to the Principal Investigator that stand apart from the issues that surround all other types of Key Personnel. These should be discussed first.

2.1 Name

2.1.1 Proposed Solution

We depend too heavily on a manual inspection process to identify that discrepancies exist, and to follow up with the applicant to make sure that the name we have on record is consistent with the name that was submitted with the original data stream. While the manual process was sufficient for pilot phase 1, it cannot be sustained if we are to expand the scope (and volume) of operations.

Therefore, the proposed solution is to restore PI name discrepancies to their original *error* severity. At the same time, we can loosen some of the stringency in the discrepancy checking, and make certain *types* of discrepancies warnings, when they do not threaten the data integrity that we are trying to maintain in the Commons. To this end, the following features are proposed for the new discrepancy-checking algorithm:

- All elements of the name (first, last, middle, prefix, suffix) shall be compared *without* case sensitivity.
- All elements of the name shall be compared without regard to embedded spaces.
- Embedded punctuation *shall* be considered when making comparisons.
- Middle name does not need to match exactly. If the middle names do not match, but it is determined that this is because the Commons or data stream is providing a middle initial only, then a second attempt will be made, using only the first character of the middle name. If there is *still* no match based on middle initial, this will be considered a fatal error condition. If the middle initials match, this will be considered a *warning* condition. In the warning message, the applicant will be alerted to the fact that the middle name that appears in the Commons differs from what has been submitted in the data stream.
- *If provided* with the data stream, the name prefix must match the Commons exactly. It is forgivable to exclude the prefix in the data stream, since this is not a critical element in a personal identification. However, if given, we must verify that it matches what is stored in the Commons, so that the applicant will not be left with a false impression that the submitted prefix has somehow been accepted and applied to his profile record. A prefix discrepancy

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

will be considered a *warning* condition. In the warning message, the applicant will be alerted to the fact that the prefix in the data stream does not match what we have on file for him in the Commons.

- These validations do not apply, if the PI name is not provided in the data stream at all. If the PI is identified by a Commons User ID alone, then this shall be treated as valid, although we should at least issue a warning to ensure that the PI pays attention to the name that results in the grant image when it is time for him/her to verify the submission in Commons. However, if *any* component of the name is supplied in addition to the Commons User ID, then *all* of the rules described above shall be applied to verify that the supplied information does not conflict with the Commons PPF.

In addition to this revamped matching algorithm, the following supportive measures are assumed for pilot phase 2:

- Our on-line electronic submission guidelines explain these expectations clearly, for all applicants.
- A Web service will be available (discussed in more detail later), which will make certain key portions of a person's profile visible to service providers.
- Service providers will be *strongly* encouraged to take advantage of the new Web service, to pre-populate the PI name fields of the electronic 398 application, and to make these fields non-updatable.

Finally, the name that appears in the grant image will be the same as it appears in the data stream, since it is more important to reflect what the user entered with the submission, than to make sure the stored image matches the profile at that moment in time. The validation already ensures that the data stream name is extremely consistent with the profile anyway, so at this point we've already verified that data integrity is not really an issue. There will be only one exception, where the grant image will *not* reflect the data stream name. For those who take the high road and identify the PI by Commons User ID alone, we will generate the name in the grant image based on the name that is on file in the Commons PPF.

2.2 Degrees

2.2.1 Proposed Solution

The proposed solution is to expand the data stream so that the face page degrees for the PI are explicitly identified by the applicant. Since it is conceivable that the Commons PPF will supply more than the requested three degrees, we cannot depend upon this information to be a simple derivation from the Commons for all cases.

Since the face page degrees were not accepted in the pilot phase 1 data stream, we currently do not have any validations proposed for these. The following validations are proposed:

- Duplicate degree designations—like Ph.D. listed twice—will be treated as an error condition.
- The face page degrees must be a subset of the Commons PPF degrees (for the PI). Some flexibility must be allowed when doing this comparison. Since an applicant may indicate

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

degree designations that are not adequately described by our established set of degree codes, we should match each face page degree against—

- *Degree Code* if the PPF degree code is not one of the “other” degree types
- *Other Degree Text* if the PPF degree code is one of the “other” degree types

As long as a match can be established for each face page degree based on one of these criteria, they will be deemed acceptable. Any degree that cannot be matched against the Commons PPF information will be treated as an error condition.

Also, the matching logic should not be case-sensitive, and should disregard any embedded spaces or punctuation.

In addition to this, the following supportive measures are assumed for pilot phase 2:

- Our on-line electronic submission guidelines explain the expectations for face page degrees clearly, for all applicants.
- A Web service will be available (discussed in more detail later), which will make certain key portions of a person’s profile visible to service providers.
- Service providers will be encouraged to take advantage of the new web service, to pre-validate degrees as they are entered, rather than waiting until a submission is attempted.

The grant image shall present the degrees which were submitted with the *data stream*.

2.3 Position Title

Position title will be validated to ensure that it matches the position title associated with *one* of the employment records for the PI, at the applicant organization.

2.4 Contact Information

2.4.1 Proposed Solution

Reinstate the *fatal* severity of this validation. All elements of the PI contact information must match what has been entered as *one of* the PI’s employment addresses (at the applicant organization). This includes:

- Street Address
- City
- State
- Postal (Zip) code
- Country
- Phone Number
- Fax Number
- Email

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

There will be no exceptions, and no special provisions, except that differences in capitalization, punctuation or white space will be ignored when comparing the data stream information to the PPF. Any discrepancy will be treated as an error condition.

As with the PI name, the validation of contact information is only applied if *any* of the contact information is supplied in the data stream. The absence of contact information shall be interpreted as an affirmation that whatever is on file in the Commons should be used. It will be useful to at least issue a warning at the point of receipt, so that the applicant is alert to verify the contact information when it comes time to verify the grant image.

Note that there is an implied relationship between the position title (discussed in the previous section) and the contact information, since these are each associated with a specific employment record within that individual's employment history. For the sake of flexibility and ease of use, we should not strictly enforce that the position title and contact information used on the application exist *together* as a combination in the Commons PPF at this time.

In conjunction with this, the following supportive measures are assumed for pilot phase 2:

- Our on-line electronic submission guidelines explain these expectations clearly, for all applicants.
- A Web service will be available (discussed in more detail later), which will make certain key portions of a person's profile visible to service providers.
- Service providers will be *strongly* encouraged to take advantage of the new Web service, to pre-populate the PI contact information fields within their electronic 398 applications, and to make these fields non-updatable.

3 Key Personnel Other Than the Principal Investigator

3.1 Relaxation of Policy

3.2 Name

Either a name or Commons User ID will be required for each key person. If the name is supplied *with* a Commons User ID, we will be obliged to compare this name with the name that appears on the identified Commons PPF record. The same matching techniques will be employed as described for the *Principal Investigator Name*, above. As a result of the name match, either an error or warning condition may result, as described earlier. If the name in the data stream matches the Commons PPF—within prescribed tolerance—it will be accepted. The name that appears in the grant image will be the same as it appears in the data stream, since it is more important to reflect what the user entered with the submission, than to make sure that the stored image matches the profile at that moment in time. The validation already ensures that the data stream name is extremely consistent with the profile anyway, so at this point we've already verified that data integrity is not really an issue. There will be only one exception, where the grant image will *not* reflect the data stream name. For those who take the high road and identify a key person by Commons User ID alone, we will generate the name in the grant image based on the name that is on file in the corresponding Commons PPF.

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

If the name is supplied but there is *no* accompanying Commons User ID, then the name will be unconditionally accepted. The only validation we will enforce is that a minimum of last and first name have been provided.

Note that this last validation requirement has a certain implication for those who wish to tease non-key staff positions into the Key Personnel section. They will have to come up with a way to express the “TBD” or generic position using as a first and last name, cognizant of the fact that the grant image generation will present the last name first, will separate the last name from the first name using an intervening comma, and will sort the remaining key personnel into ascending alphabetic order, after the PI has been listed (PI must always be listed first). So, while it may be possible to list unnamed staff positions in this section, we are still not making it particularly easy to do so.

3.3 Commons User ID

As discussed earlier, a Commons User ID is not required for any key personnel other than the PI. We will still verify that a Commons ID has been provided for the key person who has been designated with “Role on Project” of “PI”.

When supplied for any given key person, the exchange validation shall verify that the Commons User ID does actually exist in the eRA database.

It is customary, but not *required* by the data stream, to list the PI first. We will continue to have flexibility in the data stream validation requirements, and not require the data-streamed personnel to follow any particular order.

3.4 Organization

Formerly entered as the DUNS number for the organization, this will now be entered as text since the Commons account restriction no longer applies. To keep things from getting too complicated, it is acceptable to receive text even for the PI if the applicant wants to supply it. While it is fairly important that the PI name match with what appears on the face page, the organization name probably does not need to be an exact replica of what is formally presented on the face page and can be abbreviated in the Key Personnel list if the applicant so wishes.

If an organization name is excluded from the data stream for the PI, this too can be accepted, since we can still generate the organization name for the grant image using the applicant organization’s DUNS number and IPF information.

Aside from the PI, an organization name *must* be supplied for every other key person who is listed in this section.

3.5 Role on Project

This was a source of discontent among applicants during pilot phase 1. Aside from being required to have Commons accounts for all key personnel, the coded restrictions for “role on project” were also a sore spot. Since Key Personnel were being mapped to traditional PERSONS and PERSON INVOLVEMENTS entities, the role on project attribute had to be coded in accordance with the value domain that was already established in the enterprise database. Thus, all Key Personnel,

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

regardless of their specific focus and value to the project, had to be designated with the non-descriptive role of “KP”, which simply stands for “Key Person”.

The Key Personnel section will no longer map to PERSONS and PERSON INVOLVEMENTS. Therefore, we no longer need to be constrained by the coded values that are acceptable in the database for person-role-type. With the exception of the PI, the role expressed on the paper application for key personnel is fairly textual and need not be encoded in any fashion at this time. Therefore, the schema will be relaxed to accept more descriptive phrases for “role on project”, and this information will be captured in a database structure that offers more flexibility.

3.6 Grant Image

When the Key Personnel section is generated for the grant image, *all* of the elements described above shall be presented: Name, Commons User ID, Organization, and Role On Project. Presenting the Commons User ID on the image is new, but is in keeping with anticipated revisions to the PHS 398 form which are soon to be approved and published in May 2004.

The key personnel shall present the Key Person section in the order prescribed by the form instructions: PI first, and then all other staff in ascending, alphabetical last/middle/first name order.

4 Biographical Sketches

Biosketch information was a problem during pilot phase 1 due to a combination of (a) lack of instructions for applicants prior to submission, and (b) a flawed assumption that the Commons PPF would support (and thus serve as the input source for) the “Education/Training” table in the biosketch.

For pilot phase 2, the initial, structured components of the biographical sketch will continue to be accepted in the Positions/Honors/Citations attachment, even though this information is *not* related to positions, honors, or citations. However, it should be noted that the inclusion of this information in the attachment is a matter of expedience, and is not intended to stand as a permanent solution.

The ultimate solution, which we will *not* attempt to implement in pilot phase 2, will be to collect the structured information in XML components, and not rely on these to be included in a word-processor format. Ideally, *all* information would be transmitted via XML stream; however, we have recognized that this is not really feasible in all cases, and have justified that certain sections should be submitted as PDF attachments, when it can be demonstrated that there is a real chance of losing *significant* presentation value by going to plain (as opposed to rich) text format. No such justification has been demonstrated for the structured sections of the biosketch. Not only that, it would be advantageous down the road to be able to blend the Training/Education block into the Commons PPF as discrete elements, so that it might be edited in one central location, queryable by year, institution, field of study—and maintained only once instead of being submitted again and again for each research application.

But for pilot phase 2, we are still operating under the constraints of our PDF-generation software, which would make it impossible to assemble the composite biosketch for an individual without injecting a page break between the desired structured information, and the first attachment

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

(positions, honors and citations). Simply having a forced page break between the two attachments is not being received kindly right now, so adding one more page break to the current mix would not be acceptable. Therefore, we will wait to implement a structured-component approach to collecting the initial biosketch information, until more advanced grant-image generation capabilities are incorporated.

Also in pilot phase 2, we will support the submission of biosketches that are *not* associated with a key person. These will be represented by a new XML component, which will be identical in structure to the biographical sketches submitted for key personnel. Therefore, the new, non-key person biosketch component will include *only* the following:

- Positions, Honors, Citations (to include the embedded structured information from the top of the Biosketch page)
- Research Support

The only difference will be that the non-key biosketch will *not* be a subcomponent of the KeyPerson component. A new, “non-key personnel” tag will be introduced, and used to assemble all the non-key biosketches for the application.

In the following sections, the treatment of each segment of the Biographical Sketch page will be discussed separately.

4.1 Introduction Boilerplate

The boilerplate text that appears at the top of every biographical sketch shall continue to be supplied in the Positions/Honors/Citations attachment. Specifically, this refers to the “BIOGRAPHICAL SKETCH” title, horizontal lines, and the two lines of instructions which appear below the title (and precede the name and position title).

4.2 Name

This shall continue to be included in the Positions/Honors/Citations attachment, as was done in pilot phase 1.

4.3 Position Title

This shall continue to be included in the Positions/Honors/Citations attachment, as was done in pilot phase 1.

4.4 Education/Training

This shall continue to be included in the Positions/Honors/Citations attachment, as was done in pilot phase 1.

4.5 Positions, Honors, Citations

This shall continue to be received as a separate attachment, as it is received today. The attachment shall continue to include the first-page banner (with the BIOGRAPHICAL SKETCH title and two lines of instructions), the person’s name, position title, and the education/training table.

eRA/CGAP	Version: 0.1
Instructions for Collecting Person Information	Last Updated: 1/13/2004 2:06 PM

4.6 Research Support

This shall continue to be received as a separate attachment, as it is received today.

5 Web Service

As alluded to earlier, it is assumed that a Web Service will be developed and made available to service providers, allowing access to certain key elements of a PI's Commons PPF. The Web service should make available the following information:

- Full name (last, first, middle, prefix, suffix)
- Degree set (degree code, other degree text for each)
- Contact information, related to this person's employment at the applicant organization. Includes street address, city, state, and zip code, country (ISO-3166 code), phone number, fax number, and email. Multiple occurrences may be returned, depending upon the number of addresses that the PPF reflects for this particular person, at that organization.
- Position title (most recent for the applicant organization)

Input to the Web service will be the Commons User ID and the DUNS number of the applicant organization. The Commons User ID will point to the appropriate PPF entry, whereas the DUNS number will point to the pertinent employment (contact) information within the indicated PPF.