

# NIH SBIR Technical Conference Call

## Questions and Answers

**May 28, 2003**

*1. As proposed, the submission would include quite a few attachments, e.g., the Resources page is chopped into two files and the Research Plan into eight. I'm not sure why some of these divisions have been created and it seems like it will be extra work for the applicants. Can you shed any light on this for me?*

On the original form, these attachments usually correspond to sections in the form and represent distinct content that an applicant would probably compose at different times. It is assumed that an interactive application of some sort would guide the applicant in putting these attachments together and provide instructions on the content of each section.

Each section of the application has some rules on format and content (page counts, font size, etc.). Unless we have them as separate files, it is unlikely that we will be able to automate the validations while composing the application or while validating them at NIH. That would mean more manual reviews and the loss of automation benefits.

Reviewers need application sections to be linked or bookmarked. Both reviewers and NIH personnel need to be able to jump to a section top very rapidly. This is one of the major reasons we broke up the application into segments. NIH is spending a ton of money to manually bookmark every application that is currently scanned in and stored as a bookmarked PDF. With the segments as files, we can bookmark automatically when we reassemble the application file. One possible solution might be to require the preparer to provide the bookmarks. However, this model would be difficult to implement in reality.

This is a very early pilot. If, with your feedback and users' feedback, we find better solutions that meet all the requirements, we can change some things.

We are constrained by paper processes. Ideally, electronic applications should neither aid nor handicap the submitter when it comes to review time. Most of the paper rules still need to be enforced.

One more reason: All the sections for which the users have expressed the need for control of the formatting (fonts, special characters, graphics, anything but text) are requested in PDF and thereby not included in the XML schema itself.

The approach is not fixed in stone and can be negotiated after the pilot with NIH Receipt and Referral as well as with Grants.gov

*2. We derived an organizational profile from the eRA web site using the current namespace. Does the NIH staff have any plans to test or implement the organizational profile upload in the near future?*

The exchanges of personnel and institution information are part of the CGAP project. They are scheduled to be addressed in the July to October timeframe. The mandatory use of the Business Partner Network and Grants.gov may influence the solutions.

3. Are NIH staff members set up for testing transmission of XML, and, if so, are they going to start with “bare” XML documents (by email) or do they want them enclosed in SOAP from the start? (We are prepared to do either.)

We would like a complete application comprising the XML schema and the PDF attachments packaged as a SOAP with attachment message using the revised specifications that will be published the week of June 6.

In early July, we would like to have the complete round trip of ticket: acknowledgment of submission return message, application package retrieval and load into the eRA databases.

4. Is NIH ready to receive the 398 along with the attachments associated with it, or will they test the data portion first?

The NIH will receive the complete package—SOAP with attachments including the XML schema and PDF attachments.

5. What decisions are being made about formats and format checking of pdf files or other attachments? Does NIH leave the responsibility for checking proper format (margins, fonts, etc.) up to the vendor, or are they planning a scan as part of their acceptance?

The vendor should be responsible for up-front checking of the proper format. NIH is investigating the possibility of being able to validate format within PDF files. Depending on technical capabilities, NIH will do either an automatic or manual format check.

6. In working with the schema (commontypes.xsd, nihschema.xsd, raresearch.xsd), type definitions tend to be text and not codes. Do they want to follow the same example with the organization and professional profiles?

For example the following defines the Ethnicity Type in commontypes.xsd:

```
<xs:simpleType name="EthnicityType">
  <xs:restriction base="xs:token">
    <xs:enumeration value="Hispanic or Latino"/>
    <xs:enumeration value="Not Hispanic or Latino"/>
    <xs:enumeration value="Do not wish to provide"/>
  </xs:restriction>
</xs:simpleType>
```

As you can see, the enumeration values are text descriptions and not codes. There must be a reverse lookup for populating tables.

Many of the pattern validations that are found in the current copy of commontypes.xsd will be removed in the next release of the schemas, which is planned for the week ending June 6.

7. Should each institution submitting a 398 include its NIH organization profile, and does the “ProfileIdentifier” element of the “KeyPerson” element relate to the NIH professional profile?

The NIH organization profile will not be submitted as part of the PHS 398 data stream. The DUNS number that identifies the submitting organization will relate directly to the NIH eRA Commons organization profile.

The ProfileIdentifier has been replaced in the schemas by the Commons AccountIdentifier. This will relate to the profile that is maintained on the NIH eRA Commons.

8. *Are there any data requirements for the organization or professional profiles that are not contained on the NIH eRA Commons?*

No additional data requirements, beyond what is currently maintained in the NIH eRA Commons, are known at this time.

9. *Are they interested in a data dictionary of all the data GAMS collects on the institution and personnel?*

Yes, we would like to look at it from a professional interest if there are no confidentiality–proprietary issues. We do not know if we will use any of it until the analysis is done.

10. *How should we gain input about data elements we would like to see added to the schemas?*

The eRA Communications and Outreach Branch (COB) will coordinate this effort. Recommended additions and changes to schemas can be sent to Scarlett Gibb at gibbs@od.nih.gov. We will distribute them to the correct personnel and if they need to be vetted through our Change Control Board, they will be. Once a decision has been made, the COB will notify all SBIR contacts and post all changes to the eRA website.

11. *What input does the NIH staff want about security at this stage? (The recommendations in the report were adequate from our point of view.)*

All input is welcome. The more consensus we have the better off we are. For the pilot, we are going to proceed with the approach that is currently documented in our specifications. For the long-term, we will be looking at improving this approach.

**A short white paper with suggestions, technical background and suggested applications would be a good vehicle for providing input.**

12. *It looks to me like the KeyPerson element in the schema can only exist once. While it is wrapped in a sequence, the sequence has no maxOccurs field so it can only appear once. Is this the intention? And, if so, who is that KeyPerson supposed to be?*

The KeyPerson element is supposed to repeat, unbounded. We have a number of other schema changes that we are incorporating at this time.

New versions will be available by end of the week of June 6, along with a set of change notes.

13. *The Packaging*

*<[http://era.nih.gov/Docs/Packaging\\_Electronic\\_Grant\\_App\\_A\\_04-29-03.pdf](http://era.nih.gov/Docs/Packaging_Electronic_Grant_App_A_04-29-03.pdf)> an Electronic Grant Application document on the NIH Web site breaks the research body of a 398 grant proposal into seven separate parts, each to be submitted as a separate PDF. Here is a list of all the PDF documents, with the research body (parts 8-15):*

1. *Project Description*
2. *Positions, Honors and Citations for USER A*
3. *Research Support for USER A*
4. *Positions, Honors and Citations for USER B*
5. *Research Support for USER B*
6. *Facilities*

7. *Major Equipment*
8. *Introduction to Revised Application*
9. *Introduction to Supplemental Application*
10. *Research Plan sections A through D*
11. *Research Plan sections E and F*
12. *Research Plan section G*
13. *Research Plan section H*
14. *Research Plan section I*
15. *Research Plan section J*

*Will there be page numbers on the pages of these eight PDF documents? If yes, will the page numbers be sequential or will each document start at page 1? By “sequential” I mean, for example, that the first page number of Research Plan section G will be one more than the last page of Research Plan section E and F.*

Within the Research Plan sections A–D, we should try to preserve the page numbers as entered by the submitter.

It is possible that Grants.gov will provide a solution to the pagination issue.

**Any suggestions or tips are welcome, in particular tools, techniques and applications. Keep in mind that bookmarking within one document may not satisfy business rules for internal electronic grant processing. This needs more research.**

*14. The CGAP Project Plan Summary and Milestones document, dated April 10, 2003, discusses “revised specifications” for the XML application schema version 1.1 with a projected date of 04/14/03. We find only PHS 398 Technical XML Schema document dated 2/25/2003. The same thing applies to the other deliverable mentioned—can’t find current releases. Where should we find the updated documents that are mentioned in the Project Plan Summary?*

New deliverables will be supplied the week ending June 6. This will include schemas and revised packaging specification.

We will baseline these specifications, and, from that point forward, changes will be tracked in a more formal way.

*15. In the XML Technical Schema that we do have, the data element definitions for state codes do not make considerations for “provinces” (such as Canada) or some other common facets that are found in the grantee profiles that we are accustomed to seeing. How will these be handled?*

State codes in the schema are intended to accommodate Canadian provinces as well as U.S. states. Any other divisions within countries must be expressed within one of the address lines.

*16. The data element definitions for postal code and for telephone numbers provide for two configuration patterns (5-digit and 5-digit plus 4-digit) but neither of these patterns will accommodate foreign numbering systems. How will this be handled?*

We are going to relax the schema validations on these elements.

*17. How will we handle a revision prior to the deadline when a proposal has been submitted and the applicant finds they need to make a last minute change?*

The entire application will need to be submitted for revision.

Before the application deadline, the system will match for duplicates and overwrite the original submission. If the application has already been referred, we will need to send notification to the referral officer, the SRA, and the IC. After the deadline, the applicant will need to send the revised application in hard-copy form to the SRA.

In subsequent versions, a more flexible mechanism will be defined for error corrections or addendums after submission and before review.

## Notes

***Relationship between the professional profile and the CGAP biosketch***—An area that concerns me is the maintaining of the Professional Profile that is currently on the NIH eRA Commons, but is also part of the Biographic Information as part of the CGAP. For most PIs, information on their profile changes frequently (monthly or even more often). Currently the PHS 398 instructions for the Biographical Sketch limit the total length to four pages and Items A and B (together) may not exceed two of the four-page limit. A PI can choose to feature different publications depending on the content of the proposal. If there is no intent to have the publications in the Professional Profile synchronized with the CGAP, then there is no issue (these are attachments). Alternatively, if there is a need to have the information better synchronized, a decision could be made to automatically update this information during the proposal submission.

Ideally, the submission should be a subset of what is in the profile biosketch. This will be addressed when we analyze the profile interchanges.

***External documents to be attached in the NIH Biographical Sketch. the Positions Honors Citations File Identifier (Covers sections A&B), and the Research Support File Identifier (Covers section C)***—Along with the rest of the attachments, the mechanism for enforcing lengths, margins, and fonts needs to be spelled out. Are the attachments limited to PDF files or are Word or other binary documents acceptable? We assume that binary attachments will be coded with base64. Does NIH want to have the submitting organization pretest the formats? Do they have any algorithms in hand? Using a standard set of algorithms would be helpful.

For the pilot, we have limited the format to PDF. Future releases may increase the number of formats accepted. We are relying on standard SOAP protocol for the submission.

## Additional Questions

- Some additional questions were raised during the conference call regarding datatypes and validations in the schema. The following schema changes will be made in response:
  - Animal Assurance Number is currently defined as an integer. This will be changed to string.
  - The minLen facet for the CurrencyType indicates a length of four. There should be no minimum length for the currency type, so this restriction will be removed.
  - The various “name” types have pattern facets that are very restrictive and do not accommodate punctuation. These validation patterns will be removed from the schemas.
- A plan will be published in June delineating how we are going to proceed with the pilot project.
- It was suggested, and agreed, that any future plans to change the schema will be published in advance.