

UTF-8 Characters Impact

1 Current Situation

Grants.gov is currently not configured to handle UTF-8 characters resulting in transformations of some commonly used special characters (e.g. accent, tilde, umlaut characters in names, Greek letters in scientific documentation, etc.) when they are written to the file system and database. This behavior can be seen in several scenarios:

- Retrieved Grants.gov data contains “??” or garbled characters when displayed on the user interface or generated zip file PDFs/XML and when returned in the web services
- S2S submissions are rejected when the hash value calculated using the transformed special characters does not match the applicant supplied hash value
- The transformation may cause field values to exceed the maximum field length and cause validation errors, database update errors, or truncated data

2 UTF-8 Characters Implementation Impact

With the proposed implementation of UTF-8, Grants.gov will be able to store data and pass to its users as received from the PDF submission, User Interface, or S2S system. The following sections describe the impacts on Applicants and Grantors. A summary table is included in the attached Appendix.

2.1 Applicant Impacts

Since retrieved Grants.gov data will be provided as received on the user interface, in generated PDFs/XML, and by the web services:

- Before using special characters in applications, Applicants should be aware of how agencies handle special characters:
 - If the Grantor system is UTF-8 compatible, Grantors will be able to view submission data as submitted by the applicant
 - If the Grantor system is not UTF-8 compatible, Grantors may not be able to see or process submission data with special characters
- Applicants using S2S systems not compatible with UTF-8 will continue to have issues with sending and receiving special character data to and from Grants.gov

2.2 Grantor Impacts

- If Grantor systems are not compatible with UTF-8, Grantors are likely to experience similar issues as currently experienced by Grants.gov such as garbled characters and rejections due to incorrect hash value and field lengths exceeded. However, there may be additional unpredictable outcomes while special characters are processed by Grantor systems.

- Grantors may need to update their application instructions to explicitly state that applicants should not use special characters in their grant applications. However, Grants.gov will send any data received from the applicant even if it contains special characters.
- Grantors that use the Grants.gov generated flat PDF to review submissions will experience the following issue:
 - Grantors opening a PDF file with certain Asian and Middle-Eastern characters may be prompted to download and install an Adobe Reader add-on font package. If they successfully do so, foreign characters will be displayed as entered by the applicant. Otherwise, data will be missing in several fields – not just in the field with foreign characters.
 - Agency wide security settings often prevent users from installing the required Adobe Reader add-on font package. Some users might even not be prompted to download the add-on, and automatically displayed the PDF with blank fields.

3 Mitigation Strategies

- Release 14.0 will be available in the Training (AT07) environment on Tuesday December 30, 2014 to allow for early user testing.
- Expanded communications to business and technical staff of both Applicants and Grantors.

Appendix: Impact Summary Table:

Applicant UTF-8 Compatibility	Grantor UTF-8 Compatibility	Applicant Impact	Grantor Impact
PDF Applicant using UTF-8 Compatible Package	<ul style="list-style-type: none"> Grantor uses Grants.gov UI Grantor S2S with Grantor system UTF-8 Compatible 	<ul style="list-style-type: none"> In most cases, Grantors will view the application as submitted by the applicant including special characters. Until Grantors install Adobe Reader add-on font packages, certain Asian and Middle-Eastern characters may not be visible to Grantors. 	<ul style="list-style-type: none"> If applicants use certain Asian and Middle-Eastern characters, Grantors may be prompted to download and install an Adobe Reader add-on font package. Agency-wide security settings often prevent users from installing the required Adobe Reader add-on.
	<ul style="list-style-type: none"> Grantor S2S with grantor system not UTF-8 Compatible 	<ul style="list-style-type: none"> Applicant submissions with special characters should no longer be rejected by Grants.gov but may be rejected by the Grantors due to the inability of the Grantor System to process the full UTF-8 character set. Grantors are unlikely to view special characters in the submitted application. 	<ul style="list-style-type: none"> Grantors are likely to experience current Grant.gov incompatibility issues including garbled characters and rejections due to incorrect hash values and field lengths exceeded. Additional unpredictable outcomes may also be experienced by the Grantor Systems while processing special characters. Grantors may need to update application instructions to explicitly state applicants should not use special characters in their grant applications.
Applicant S2S with applicant system not UTF-8 Compatible	<ul style="list-style-type: none"> Grantor uses Grants.gov UI Grantor S2S with Grantor system UTF-8 Compatible Grantor S2S with grantor system not UTF-8 Compatible 	<ul style="list-style-type: none"> Grants.gov and the Grantor systems will receive and process the application as submitted by the Applicant system. However, special characters in the UTF-8 character set not supported by the applicant system will not be sent to Grants.gov or the Grantors. 	None

Applicant UTF-8 Compatibility	Grantor UTF-8 Compatibility	Applicant Impact	Grantor Impact
Applicant S2S with applicant system UTF-8 Compatible	<ul style="list-style-type: none"> Grantor uses Grants.gov UI Grantor S2S with Grantor system UTF-8 Compatible 	<ul style="list-style-type: none"> In most cases, Grantors will view the application as submitted by the applicant including special characters. Until Grantors install Adobe Reader add-on font packages, certain Asian and Middle-Eastern characters may not be visible to Grantors. 	<ul style="list-style-type: none"> If applicants use certain Asian and Middle-Eastern characters, Grantors may be prompted to download and install an Adobe Reader add-on font package. Agency-wide security settings often prevent users from installing the required Adobe Reader add-on.
	<ul style="list-style-type: none"> Grantor S2S with grantor system not UTF-8 Compatible 	<ul style="list-style-type: none"> Applicant submissions with special characters should no longer be rejected by Grants.gov but may be rejected by the Grantors due to the inability of the Grantor System to process the full UTF-8 character set. Grantors are unlikely to view special characters in the submitted application. 	<ul style="list-style-type: none"> Grantors are likely to experience current Grant.gov incompatibility issues including garbled characters and rejections due to incorrect hash values and field lengths exceeded. Additional unpredictable outcomes may also be experienced by the Grantor Systems while processing special characters. Grantors may need to update application instructions to explicitly state applicants should not use special characters in their grant applications.