Overview

The RFC Editor intends to assign a Digital Object Identifier (DOI) to each RFC in the series. https://datatracker.ietf.org/doc/draft-iab-doi describes the motivation for and expected use of DOIs in the series. As that document states, assigning DOIs entails extracting information from the series to date to be used as input to a bulk registration for existing RFCs, and changes to the production system to automatically acquire and incorporate DOIs into future RFCs.

Deliverables/Tasks

The developer will work with the RFC Editors' programming staff to

- Identify the data needed to apply for a DOI for an RFC, and for each data element, identify the appropriate source for that data.
- Build a program to extract that data from the existing series necessary to apply for a DOI for each existing RFC.
- Construct any additional needed automation in making a bulk application for those DOIs
- Integrate the automated acquisition of DOIs for future RFCs into the workflow for producing that RFC
- Construct any needed automation for updating the existing index, info pages, and the XML citation library to include the assigned DOIs
- Locate existing DOIs for the documents from other organizations that are in our XML citation libraries (particularly bibxml2) and update the entry in that library.

Detailed Description

The International DOI Foundation (IDF, www.doi.org) is the governance and management body for the federation of Registration Agencies (www.doi.org/registration_agencies.html) providing DOI services and registration. CrossRef (www.crossref.org) is a highly regarded registration agency.

CrossRef provides several mechanisms for acquiring a DOI for a document. An overview of the mechanics is available at http://help.crossref.org/#how_crossref_works. This project will likely use the HTTP interface described at http://help.crossref.org/#using_http_to_post for both the bulk assignment of DOIs for the existing documents in the series, and for automating assignment of DOIs for future documents. The specific design of these acquisition steps will be created with the RFC Editor's programming staff during the project.

The HTTP query described above requires an XML document containing specific information about the documents being registered. The required information is

described at http://help.crossref.org/#elements$Report-p and the XML can be verified before submission as described at http://help.crossref.org/#verifying_your_xml. This project will automate extracting the required information from the existing series and building the required XML documents for submission, ensuring that the produced document structure satisfies the above verification tests. Details on the necessary information and the required format are at the CrossRef schema for standards and report papers. Standards-track RFCs (Proposed Standard, Draft Standard, and Internet Standard) and BCP RFCs will be processed using the standards schema. All other RFC types will be processed using the report papers schema. It is anticipated that most, if not all, of the necessary information will be extracted from the rfc-index. For future documents, the necessary information will come from the same data source that the entry in the rfc-index will be constructed from.

The project will automate a bulk update to the following products to reflect the DOI information for existing RFCs.

- http://www.rfc-editor.org/rfc-ref.txt
- The bibxml products
- The individual RFC information pages (e.g. http://www.rfc-editor.org/info/rfc6635)
- rfc-index.xml
- rfc-index.txt
- rfc-index.html and rfc-index2.html

The project will also automate the addition of the acquired DOI for each new RFC into each of the above products.

Acquiring DOI for other series, or for subseries of the RFC series, is not in scope for this project, but any development should anticipate those efforts, building reusable components where possible.

The grammar for rfc-index.xml (see http://www.rfc-editor.org/in-notes/rfc-index.xsd) will be updated to allow a <doi> element as a child of <rfc-entry>. Specifically:

```
<xsd:element name="doi" type="xsd:string" minOccurs="0"/>
```

will be added to the end of the sequence listed for the rfc-entry element.

A resulting rfc-index.xml will look like

```
<rfc-entry>
   <doc-id>RFC6949</doc-id>
   <title>RFC Series Format Requirements and Future Development</title>
     …
   <stream>IAB</stream>
   <doi>(the acquired DOI for RFC6949)</doi>
</rfc-entry>
```

The developer will work with the RFC Editor to determine the exact representation of DOIs in rfc-index.txt, rfc-index.html, and rfc-index2.html

The bibxml products are currently at http://xml2rfc.ietf.org/public/rfc/. A typical RFC entry there currently contains:

```
<?xml version='1.0' encoding='UTF-8'?>


<reference anchor='RFC6949'>


<front>

<title>RFC Series Format Requirements and Future Development</title>

<author initials='H.' surname='Flanagan' fullname='H. Flanagan'>

<organization /></author>

<author initials='N.' surname='Brownlee' fullname='N. Brownlee'>

<organization /></author>

<date year='2013' month='May' />

<abstract>

<t>This document describes the current requirements and requests for enhancements for the format of the canonical version of RFCs.  Terms are defined to help clarify exactly which stages of document production are under discussion for format changes.  The requirements described in this document will determine what changes will be made to RFC format.  This document updates RFC 2223.</t></abstract></front>


<seriesInfo name='RFC' value='6949' />

<format type='TXT' octets='29181' target='http://www.rfc-editor.org/rfc/rfc6949.txt' />

</reference>
```

This project will add an additional seriesInfo element:

```
<seriesInfo name='DOI' value='(the acquired DOI for RFC694)'/>
```

Expected Development Processes and Practices

The contractor will adhere to the requirements at http://trac.tools.ietf.org/tools/ietfdb/wiki/ContractorInstructions?version=11