Language engine for including data in Rmarkdown

David M. Kaplan

3/27/2020

Overview

This document contains a proof of concept for adding data engines to Rmarkdown that would allow placing data directly inside Rmarkdown documents to create completely self-contained Rmarkdown documents. The implementation is based on the idea that data chunks will contain the contents of the data files, potentially encoded as text using some encoding method.

Though many types of data chunks are imaginable, I have currently implemented three: text, csv, RDS and a generic type where the format chunk option should be a loader function for loading in a file containing the decoded chunk data. For binary data chunks (e.g., RDS), the data must be encoded as text before inclusion in the chunk. Two different encodings are currently implemented: base64 and gpg. If the encoding chunk option is not supplied, then base64 is assumed, except for text and csv for which no encoding (i.e., asis) is assumed.

I have included helper functions for simplifying encoding and decoding content using the methods described above. Encoding and decoding operations rely on the base64enc and gpg R packages.

The implementation is quite simple, but it works well. One potential drawback is that this could encourage creating Rmarkdown documents with large datasets inside that are unreadable. This could be prevented or discouraged by limiting the size of data chunks, perhaps with an chunk option to allow the user to violate size limits if they really want to.

Implementation of language engine

```
# Helper function to decode encoded text
# Returns name of file where decoded text was saved
data_decode = function(code,encoding,options,file=tempfile()) {
    switch(
        encoding,
        asis = writeLines(code,file),
        base64 = writeBin(base64enc::base64decode(code),file),
        gpg = {
            tf = tempfile()
            writeLines(code,tf)
            on.exit(file.remove(tf))
            writeBin(gpg::gpg_decrypt(tf,as_text=FALSE),file)
        },
        stop("Uknown encoding: ",encoding)
        return(file)
```

```
# Helper function to encode (typically) binary files for inclusion in data chunks
# Silently returns encoded text. Encoded text is also spit out to console for
# copy-n-paste to Rmarkdown document. This can be surpressed by saving encoded text
# to a file using output argument
data_encode = function(file,encoding,options=list(base64.linewidth=64),output=NULL) {
  code = switch(
    encoding,
   asis = readLines(file),
   base64 = base64enc::base64encode(file,linewidth=options$base64.linewidth),
   gpg = {
     if (is.null(options$receiver))
        stop("Missing GPG receiver. See ?gpg::gpg_encrypt for details.")
     gpg::gpg_encrypt(file,options$receiver,options$signer)
   },
    stop("Uknown encoding: ",encoding)
  if(is.null(output)) {
    cat(code,sep="\n")
  } else {
    writeLines(code,output)
  invisible(code)
}
# Data engine itself
eng_data = function(options) {
 vn = options$output.var
  if (is.null(vn))
    stop("output.var must be supplied in data chunks.")
  format = options$format
  if (is.null(format))
   format = 'text'
  encoding = options$encoding
  if (is.null(encoding)) {
   encoding = switch(
     format,
     text = 'asis',
      csv = 'asis',
     RDS = 'base64',
      'base64'
   )
  }
  encoding.ops = options$encoding.ops
  if (is.null(encoding.ops))
    encoding.ops = list()
  if (!is.list(encoding.ops))
```

```
stop("encoding.ops should be a list. Got object of class ",class(encoding.ops)[1])
  format.ops = options$format.ops
  if (is.null(format.ops))
    format.ops = list()
  if (!is.list(format.ops))
    stop("format.ops should be a list. Got object of class ",class(format.ops)[1])
  # If format='text', just spit back code, possibly after splitting using a separator
  if (is.character(format) && format=='text') {
   code = options$code
   if (!is.null(format.ops$sep))
      code = do.call(c,strsplit(code,format.ops$sep))
   assign(vn, code, envir = knitr::knit_global())
   return(knitr::engine_output(options,options$code,''))
  }
  # In all other cases, decode data first
  fn = data_decode(options$code,encoding,options=encoding.ops)
  on.exit(file.remove(fn))
  # If character format given, convert to a loader function
  if (is.character(format)) {
   format = switch(
     format,
     csv = read.csv,
     RDS = readRDS,
      get(format,mode="function") # Attempt format as function name
   )
  }
  # Load data
  data = do.call(format,c(file=fn,format.ops))
  # Assign to output.var and exit
  assign(vn, data, envir = knitr::knit_global())
  knitr::engine_output(options,options$code,'')
}
# Add to knitr's list of data engines
knitr::knit_engines$set(data=eng_data)
```

Test of text chunk

Note that format="text" is optional as data chunks default to text if no format is specified.

```
This is a text.

It has two lines.

to
```

With separator

This is a text.

```
It has two lines.
t00

## [1] "This" "is" "a" "text." "It" "has" "two" "lines."

For numeric input

1,2,3,4,5,6
7,8,9
t000

## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9"
as.numeric(t000)
```

Test of CSV chunk

[1] 1 2 3 4 5 6 7 8 9

```
id,res
1,a
2,b
3,c
t1
```

 $\begin{array}{ccc} \operatorname{id} & \operatorname{res} \\ \hline 1 & \operatorname{a} \\ 2 & \operatorname{b} \\ 3 & \operatorname{c} \end{array}$

Test of RDS chunk with base64 encoding

 $\label{thm:constraint} $$H4sIAAAAAAA4vgYmBgYGZgZgNiViCTgTU0xE3XgoGBSRjIYQLid1CakYGFgRNI8yXn5xYkJpfEZ+aVFKcWosmyJCUWp0LFeMHiEPofSCfIKgcVBjCw/4BKQ9UIIJnFnJiUDDQR2XjWvMTc1GKoOiaoIGMijJEEMgUAP3IVTdMAAAA=$

```
names(t2)
## [1] "a" "b"
t2
## $a
## [1] 1 2 3 4 5 6 7 8 9 10
##
## $b
## [1] "abc"
```

Test of GPG encrypted RDS chunk

Import private key for decrypting chunk

Normally, the decryption key would not be included in the Rmarkdown document, but I am doing so here so that code works for all. This will import the test key into your GPG key ring. This test key does not have a password, but most real keys would and the keyring management software would ask you for that password when knitting the document.

First the code for generating and exporting the key. Not evaluated, but keeping around for reference:

```
gpg::gpg_keygen("Test Key","test@test.org")
id = gpg::gpg_list_keys("test@test.org")$id
gpg::gpg_export(id,secret = TRUE)
```

Next the key itself in a text chunk.

----BEGIN PGP PRIVATE KEY BLOCK----

lQVYBF6A3p8BDADcaf7tveXZUpi0IfEpmYrPP8/OSXSh3iBkd5bdTvbq/FwLGIsD dp/dFqAWS+0BqCIMFAtV63FU0G4kXYpkajdl2QU1Hy0aY9F9K0imc5JUM1SEry5F CckjzDFp3u4pmmCPWKF2jVnaHzahJfKz9J9qD9BfBSynfyQU2XgsrRqNgiqeNcOi f0674hpReawnecBwhENKMWL3801a0tP1IDx9cF16busiiOaIHIYYW6qbv178offy OOWogstsQ3EJQbPBPkkgVTn8wwGUtoorc/2AonSoz99QC4nMWbBaDUGuE9032yRv Q7Pe6bWVBuIeV5ASAfSSEypzNHB576BF6MTy+lJvhfXI41Yu97geQJM0CplJ8xav xAhIvrKjkDoW3zwrZlG54G2TidwEyXoDx7cyRVnCf9tsBCmhEDiKvzlg2IE9Fo65 +LWrD12qCKi7cu4XE28q4zy7S4adhUCBcuf1Z8wKMVvbZRXvqnAHBAK8gQxMqHMc EjWAb7rvmN9bkTUAEQEAAQAL/if4vPeGYaGIvhKkuSRvK0Iu0104tIMKUluF6IEX 6eVxgIuulr85CwLAMKX6f0+4+vuvwuKBARth5G+J2ygcrxE0SyJ4FejcQ0hsyg8N lHLaoDAzyLNSc/ye8jMd75jx2yMD0rw6JBpPYMvWou4JpcNJP000f6ucfgGd8pI/ jjotaecpHuJgLfoapeUyqIq8JK8C/WT+EdGfCpw7YObqQq4I6ZCZPuETbKMwcQOH yqfWC7bK9Lk/MvbdSWDH1j7Of/t1KaUEBZ2z5xTALqxaFgbwXh+7FybzV+09Sxsn 15deeubEQXwkbPthapjRpvRo197tJRHLJ8wQVCwag39ip5cvuWQIsej3qILKTepz VBdgZa4hIyLX8uUCAtLrVYwvWzV1oWxPLAkXJ6KPCzB0jQb7q7UUyrB0Uaavdnt2 aWBz5EuXPTaMqnzWqEKIazcXqiCSNjIEv7HWcU734IGUazYper3poYg0WYYIdUes +xbdWP/j6313N3u4a9BSd3PMvQYA4CLwr+gBfX+dybX3jq3ldB3HJS/Lv90e64rh BarRu+ByyE05BcVJZ+ZEU0cBjF/pvG1qI9mfqBuZX/e2aW11mMsxcXN1WRu5b5vE geoRwqPMNIo4JIo2hByHZeEPQLcYW/QRy5xkoNbl+udPuS3PMEUnfnPeQKursY71 ao7ZoOTUeFRemEgkvxZpFXfT+IMs9DGI/Wi6POOChSJ/Cu/QixgKOeJFUroNCyvl bW+xy0GSB325wky0M5xIny681KtvBgD7v5V6n0P2UucxZYU5hhdWaaTf5aF83vtE o88gSU5NRO1/wPFb+AFP3fw8TNtrvR1A/OakwjL+GbfhioAJ4mtPbdGUojFIAU6X czMHbaYyNwZTMImBW9uc2gDqta801HiSwC7fXnTxVoSz3E/TD6dbAnFyf1FYNntJ PLKS9H82idCqO0nrU3LtdKJx9VHJ6wLOT16D6zZAdgNBOwK9dzStayfIqQzN/FAz O1uOehX4SDRCxxgukdR4ZyeZJfdmC5sF+wZ/2mW4Tp7v3kutNAytk4JtMvLIhe2r BQkYw5eUFMq7tUqXgsXMjA0pVplUSosZknCIpoyoEU7rvS9BF9xdcpRixU5kxeYY knQg5jtb+vx3StppOvbuvFFaGgEJhNP6Tg3al7gBCOwEEAJmSTko4cyf1e45pIMF +jGbIeozSjeKPWjdJCr4q05tvKgsiAe7BulgUlNhS6Ty5JyQHsiM/WZTPko2BsN2 8Apa/nuOvYwRwFLGGXVVWV3jQroPI9Hbft9ctBhUZXN0IEtleSA8dGVzdEB0ZXN0 Lm9yZz6JAdQEEwEKAD4WIQTv16O3A7Tx/z8NeR/qzEhnRW4g7QUCXoDenwIbAwUJ A8JnAAULCQgHAgYVCgkICwIEFgIDAQIeAQIXgAAKCRDqzEhnRW4g7WxrC/94WT6J HEEgyb9Bskm2ik+c/qUW8w7JgizYRi6jqi8+qiIesh99MZ/XPm5mgMTIvKrOz/IG xaU+RKYFF5DqsAc4obg/ZmClOSY9FgDWlMEm7hEqourQxfJZXGWRNcU6DTr2tC/K GpTNkhR802LnjUePeVJU5MMuJ8eyQV+NgGhwXTIcPA6ERwHIC1n24N3QDFNoijcc pTi5p9+N33w8fBC5ZMeZwrWI6mCJjEWVbxG2zcsIJ2t7htWRM7W1rKi5lHRpQdn/ cd9WtbdDFj7ywGPnjMB2vxYVJreENGbE/LZIZPaJKJHPReWQ+GBSGkyY7nrT32SP R+qj5g00Bez7F+61EDU+SXP9PJ8fyTGtUWfTsgz+fTj2TDn39y0tL1wuSciEOAjD uia+L5qiKE9GK6mBQv78yfzZ/Z0EdJn9ZNRWs8kvs/aG9BygYMdJM5T4vvk2DcWd m061EGTg/AVUFpMuTon9tb+RCIFfVjSzat8LWcf4Me2nJeFZu+1W/1CmxkedBVgE XoDenwEMANPff6PrZirginP4HNK7g3ANmB3bDKCI1msAQspXMzvhtMcOHn8DpM+r wPUuoOo4hnYwkGHSNZ4dulrtW99mlzQWcFwDuOsvPAqc/OuEIEoOBBvc5HcpNk4d z94Vno+Dq904VnlStf6DXpGbBFZkZBoC4XVwFUSoEjD1i967ckjFUh0xE5ynlcMb 8mpS65iml4JFd572bcuo9exJ1g7IhdgFIFoDDD2eJkxEhmEHNiVd8B9/j1GHxDCq v/DOHNbgKuFk8WJUMYvupdqA30wAc5Ujnf+nURfNejgZTOiGXm5FZBrw/dha7yTP /mlnNFMBKUEBrxYyPo2JVSsYfPf1WzLL1dmv8JPC5fyEKYhEC+zBvlytRWqkZV88 DumgVEdhEnnMEVlofyF8KoVMmWYA9w/FUUKiNymZlK1PEGecqliEhXh+KEO3ncHh AyEoOZcdh5sSxUW5fNsQb+tpOfqFBs7Yye432w6ID3ZIONrnWrQ6MewWwxeAGMam x03jgyMlCwARAQABAAv9EJ0e8iicS1JuKOfUwsWHafr26ahqlhAE2EEd+6XY06JA PbqdhZIwkORBjjhIz/T8vjnSqIkGQU7NdSHVqW/u/VuhFeYIOxBSIfbrckBbE9Z+ V/z7QUjPBFMcIKsLUu+dQ2yOg1b0BHAis0I3ldqrasq9CStvz4FqY8JtZFrIfGJU rEyfYBJYEQOY/7Ne3Ap8KO/vkFx8gZLPLecgTOp2bFkCj2xbwl0rXaGl8+fP3CBA mweyok8GGFbbVDagKE1NiukpEVzHsoMyMfPkxdIMLSj0F2GzQSnhyhyGomNstuTT EC/i3/u7M9TRvLkpNTP3I6z5VNjayrp0NBs0z3sb1wNzrACELWbTtb/Lo5BVVD9Y mOMQtDi8+SKzTHci2Adpvewxnh04IiS/aXYYGcPwmEX4YdlZeV0J5mRXNsvWxYZk ${\tt HHFkbfgUkiFSF0mb9uyPD0NMldJoLXbv9+LFiU1okglietVcKK7Fyt5xCKcxbt08}$ kdYJTuWonsWeyC8tz1WBBgDcq6doxs3aFSVeLcZO//WHif+iBY1LFoexmw4irx8e LnZilDJ5i4mwcu6Q5qxao3UEyeUC7ff//Qn846TQMDDRcC3xtrbqAqVyYBE7u9EI OMyyCfosk8nNmVBpNdnsFm761UyG8GiuT6b0j8BiQTRPmH4Xlh3pSiihyuTJIVhX Y663wV8EwT9IRnYCoVqw9s5qZqJGkI4rxnABuyJui4BpmkrLry70t1xb6MdX2BPD eK5u0YJ24AmxPW5YGvXn00sGAPXLRfarrI9IgSz28+QpfYtt0Ibjp3n3AxB3ImHo oK+CLsc1vHtsdEV8hE1Wo9k5EqcdlhPBbeC6IILFqT69Ldx8jK85hxR0bYs2NVLC qyWo1T3bovPePCEenN4++VPBtVBkEt51MByNIKwC3BwOzvHcygLcHE3iXRQ40dhq AZWrPlOqwnC8x9+UqZoWCp/JRWD5qBjD6EPVAxwbtcUdjD0hZ1y51xbUaX59Vlul BGLse/OQ47m71HrF+d9rGUnlQQYAkDQsdbzijmB/tVzcRXJWbZVgjwLciofxVpoM TEYyw8+oSYDI1L3Dikejp3XymVr+9pKGmPZjLqL9Q01J9epeHt5wgLjuWTXtkVLW kbnt7vTy257BIsHGDwiJzMI7PujTlQ4B1ZTPz2WyUJ7gn1f+J9wYpN0r7qeE2pg6 cOeiPQmT5h88jWTUH/eAJOnAWx46kwgQY4uZz7xsFtCcwQgqVe9bD5MNv/bBUdPW RkF8ZbRCPRk4V12DYM/rXC2VGCFZ6OeJAbYEGAEKACAWIQTv16O3A7Tx/z8NeR/q zEhnRW4g7QUCXoDenwIbDAAKCRDqzEhnRW4g7ZayC/954y+kfmjtIzSRDBRpOo2s npOOwy7RLdOdWvab6jVecyqYsDyd/fiCXVKxALOVR31WTefOOiFSLHQactwFxQyJ zY6Y08tGkvYEXXYJR505MNzjlhNMndBqGIbKe9tA2BFLDD/6mmvMD/i9k+IhHzFT NhoczB5rE9oaApMZhAj9u9Uv2zyOosfcOPcy+RN9b2noodVS/7Ei2BjWl+V/MGqa I8oBM/ETIW/jcq+OuE8oSqoByFtFHh1DgOzOFugCWApOmAjLQwQCmDiYYtKN1GWq 11E+txLud78ZBsJQL/78MXO9V2T2dCbcIAOvOfACuoPApfu6seREOSLeImgoRg+8 7aX6HtiRXRjExDS26YNbGYzAvVTl3Zy1VptXOMwkh5CcIgtTcDv32pLWC3xvNydG P4xDMM+BVuDi6QTcFfbPtqYbuuT40FyyaSzeeOoWxvKoX2pL81VnMwvb7Uy47Dxf Ng9Af4cf3nf9UzesAVbSy1gtvlZIyX0HwtZNVLNJSS4=

```
=C6UF
----END PGP PRIVATE KEY BLOCK----
tf = tempfile()
writeLines(key,tf)
gpg::gpg_import(tf)
##
        found
                 imported
                             secrets signatures
                                                     revoked
##
            3
                        1
                                               0
                                                           0
file.remove(tf)
```

[1] TRUE

Data chunk

```
----BEGIN PGP MESSAGE-----
Version: GnuPG v2
```

```
----END PGP MESSAGE----

names(t3)

## [1] "a" "b"

t3

## $a

## [1] 1 2 3 4 5 6 7 8 9 10

##

## $b

## [1] "abc"
```

Encoding a binary file for incorporation in Rmarkdown

Below I demonstrate how to use the helper function data_encode to encode binary files for insertion in data chunks.

```
saveRDS(t2,"test.RDS")
```

base64

=WRy/

```
data_encode("test.RDS","base64")

## H4sIAAAAAAA4vgYmBgYGZgZgNiViCTgTUOxE3XgoGBSRjIYQLid1CakYGFgRNI
## 8yXn5xYkJpfEZ+aVFKcWosmyJCUWpOLFeMHiEPofSCfIKgcVBjCw/4BKQ9UIIJnF
## nJiUDDQR2XjWvMTc1GKoOiaoIGMijJEEMgUAP3IVTdMAAAA=
```

GPG

```
id = gpg::gpg_list_keys("test@test.org")$id
data_encode("test.RDS","gpg",options=list(receiver=id))
```

```
## ----BEGIN PGP MESSAGE----
## Version: GnuPG v2
##
## hQGMA9TPonHna5j3AQv/Xv7T0By1U6XFJ8TXCDK6WH11C23TCh14/bbU0BlseWvs
## McGt0Q8QUaVNJasaXwU9O3thIAN1/VML5jaE61AG1LrsyOCtc5bC9FDSO+6o4jU7
## JIeQe3A06p6abNyyCr3oheTNKkqwbmv+a7ABMNARrgRlOGM8jI1+vHQGJ5mS29pT
## r9/BGU6JzW+61j0fuN8CVEe04Qz25pc4GmYdUXberxBzNmtK5f352983a13CcmK5
## 0o5NeHIXrjS/65zSVamNGr20zXIJ06C2EarfCK4IW5XfbtymrumNnE99LFsuGUpv
## cV1ZZycgcEpR/oLGvHvdr5yBragIFpHFo/pBj6aVaszbeCbxRQS8zLw4/RKaBBXB
## 7VYDuek+jlv2B4pQqZYjtvK//OzaOVb76c67P9WoC3uPL6ivLhtmbfc2RqY+NB3u
## BvMcKXwN2johyiE4DDlCiDKAm+4Q3Ac42ECqodGubvneTB2JzdAnxQUYhRH4vkqy
## 4cGwpSU+UQwU1IsLBKBU0r8BYvgfkMo+eA42RFySpq346KjN3Laf3yzrb1h1rXTx
## UoYvNeSumpuf3n9fgjzMTo7kaZ73vQBYEZZbh3fpVKujUQ6mn8I+NiPSVMC9hiGu
## 7LNW3qQbSTZmyJ9yXEYI+qmaGJ3Npefo3alVJpKKt6p1dezWJ46EnkOaW0zC3xDc
## gjtVkw3QsNW0+bQSAMmjZgMWYKmV/MRGB3NeNehwMm6nCAqV+E6lTuKE1hporrV1
## VY5BLC6xI5pQUX6m8hiCBg==
## = gPMx
## ----END PGP MESSAGE----
```

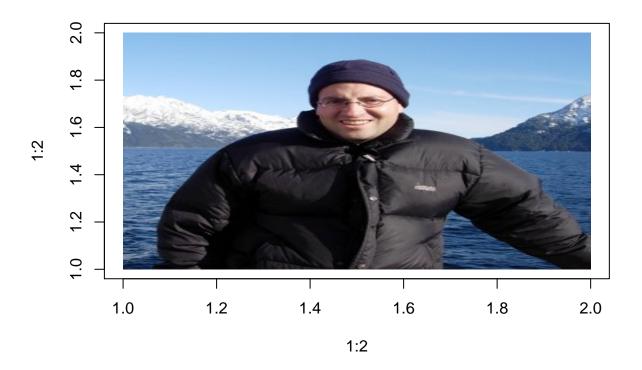
GPG saving to file

```
id = gpg::gpg_list_keys("test@test.org")$id
data_encode("test.RDS", "gpg", options=list(receiver=id), output="test.gpg")
cat(readLines("test.gpg"),sep="\n")
## ----BEGIN PGP MESSAGE----
## Version: GnuPG v2
##
## hQGMA9TPonHna5j3AQwAk7ecTP/nfXt9HUDh4kMKKpIz14hcw5Yj2jm7Jqhx+rHj
## 40df4tIisbC3qmGqUuyoNxpJw+ZNmK0IOR3inlkRiweZ5SKmGRaU3q46Qz/hZPVg
## ZQ6IOV+ijJTeGnR8nR2I4rbb7R4N/xiRiOdhKB2iqZtmE1DCjyYq609cwp6h2U2U
## OrRhPsAKmIKadjsOTy3zS+z1dnRhMNDQZvG+6Urvy62vyMI4GD/5fKfsd8hZFezU
## k4xP1xsEL58Ppd5Dsoxbrr+F0HunyCn4k4wbp8hcGlSEK1PHiT+F5gCG8L3U+nWc
## IRmdwhooXpIaLJdR8MG79d8cRBAqoCvv2bCEx616g5wVnb5uvp7bBHiqAqgoO+Ou
## cT/KNlNCkYk5h4RCN+BvwQN1DZAcOODtOiVIh1oj12Q6uQojBFjbdhROOBbJFZkZ
## mtl6jJXZkvJ4gNnt000co3Botr4VpugKRIIsUZH/jH6M7LWxKedXckkx6BmbpAwB
## x6uQpgCBfTOks/KCTZYNOr8Biq3h6+jOo20W9YJaWTtBT1jW1tl09+6JimjMHQEB
## nE02A0ZHKxv1Ka5lXjrWe/izJkHGvKZk3jwjvRJz8JF2pNxz7gHkJ9TrbcGQ79/8
## krB7hDrDQCSGuQ0iVsr0s2E6VSPjNd3NbCLJpD90x34tCDgGSgjXr4jC/5X4vYIP
## qh+xEPybQyHHZa7osEWC+Xc/A4pKW/KySP0/05aSlm3FVRpea0W1zmWXzfKl7+xR
## iXq91/dJLHVtjyz2mjWUaQ==
## =lmoB
## ----END PGP MESSAGE----
#file.remove("test.gpg")
```

Test of arbitrary binary format with loader function

Note that echo=FALSE is quite important here. Perhaps I should force this on all data chunks?

```
plot(1:2,1:2,type="n")
rasterImage(img,1,1,2,2)
```



Remove imported key

I am cleaning up by removing the imported GPG key.

```
id = gpg::gpg_list_keys("test@test.org")$id
gpg::gpg_delete(id,secret=TRUE)
```

[1] "EACC4867456E20ED"