

```
#' @title Bind to file: Load XML file and return as data.frame()
#' @param x xml file name (full path)
bind_to_file <- function(x) {

is.NullOb <- function(x) (is.null(x) | length(x) == 0) | all(sapply(x, is.null))

## Recursively step down into list, removing all such objects
rmNullObs <- function(x) {
# Replace missing elements with 'None', this is to work with Neo4j more effectively.
if (is.NullOb(x)) {
return('None')
}
out <- lapply(x, function(x) {
if (is.list(x)) {
z <- rmNullObs(x)
} else { z <- x }
z <- unlist(z)
names(z) <- names(x)
return(z)})

```

```
return(out)
```

```
}

clean_file <- try(xml2::read_xml(x) %>%
xml2::as_list() %>%
magrittr::extract2('Award'), silent = TRUE)

if('try-error' %in% class(clean_file)) {
return(NULL)
}

file <- clean_file %>%
rmNullObs %>%
purrr::map(function(z) as.data.frame(rbind(z), stringsAsFactors = FALSE))

names(file) <- names(clean_file)

newRows <- list()

for (i in 1:length(names(file))) {
```

```
newRows[[i]] <- file[[i]]="" %="">% data.frame()
colnames(newRows[[i]]) <- paste0(names(file)[i], '.', colnames(newRows[[i]]))
```

```
if (length(grep("V1", colnames(newRows[[i]]))>0)) {
  colnames(newRows[[i]]) <- names(file)[i]
}
```

```
}
```

```
# These column names are right:
```

```
output <- newrows="" %="">%
```

```
Reduce(function(dtf1, dtf2) merge(dtf1, dtf2, all = TRUE), .) %>%
```

```
purrr::map(function(x){x[is.na(x)] <- x[!is.na(x)][1];="" return(x)}="" %="">%
```

```
as.data.frame()
```

```
tester <- data.frame(AwardTitle = NA,
```

```
AwardEffectiveDate = NA,
```

```
AwardExpirationDate = NA,
```

```
AwardAmount = NA,
```

```
AwardInstrument = NA,
```

```
Organization.Code = NA,
```

```
Organization.Directorates = NA,
```

```
Organization.Division = NA,
```

```
ProgramOfficer.SignBlockName = NA,
```

```
AbstractNarration = NA,
```

```
MinAmdLetterDate = NA,
```

```
MaxAmdLetterDate = NA,
```

```
ARRAAmount = NA,
```

```
AwardID = NA,
```

```
Investigator.EmailAddress = NA,
```

```
Investigator.FirstName = NA,
```

```
Investigator.LastName = NA,
```

```
Investigator.StartDate = NA,
```

```
Investigator.EndDate = NA,
```

```
Investigator.RoleCode = NA,
```

```
Institution.Name = NA,
```

```
Institution.CityName = NA,
```

```
Institution.ZipCode = NA,
```

```
Institution.PhoneNumber = NA,
```

```
Institution.StreetAddress = NA,
```

```
Institution.CountryName = NA,
```

```
Institution.StateName = NA,  
Institution.StateCode = NA,  
ProgramElement.Code = NA,  
ProgramElement.Text = NA,  
ProgramReference.Code = NA,  
ProgramReference.Text = NA,  
FoalInformation.Name = NA,  
FoalInformation.Code = NA,  
FoalInformation.Text = NA,  
stringsAsFactors = FALSE)
```

```
full_output <- bind_rows(tester, output)
```

```
if(any(colnames(full_output) %in% tester)) {  
  warning(paste0('File ', x, ' had extra column names.'))  
  full_output <- full_output[, colnames(tester)]  
}
```

```
full_output <- apply(full_output, MARGIN = 2, FUN = function(x) gsub(" ", "", x))  
data.frame(stringsAsFactors = FALSE)
```

```
return(full_output[-1,])  
}
```