

## EII 7446 Project Update 2

Daniela Jofré Quiroz, Richard Lobos Araya, Mario Ríos Serrano, Sergio Sánchez Inostroza

En esta actualización de avance del proyecto, se logró leer la base de datos del Challenge en R. Luego de una serie de complicaciones al momento de usar MatLab y R para poder acceder a los datos estructurados de las obras de arte repartidos en 112.039 archivos (uno por cada obra) XML en la carpeta **rijksxml.tgz**, se pudo exportar con éxito. La extracción de los atributos necesarios y recorrido de los archivos en la carpeta se realizó en MatLab, generando un archivo **.mat**, con 112.039 filas, correspondientes a cada una de las obras de arte; y 6 columnas correspondientes a los atributos necesarios para realizar el Challenge. Estos atributos son **ID**: identificador de la obra; **Título**: título de la obra; **Artista**: autor de la obra; **Descripción**: breve descripción de características de la obra; **Tipo**: que tipo de obra es (pintura, dibujo, cerámica) y finalmente **Fecha**: data de la creación de la obra.

Respecto a la estructuración de la data, existen algunos vacíos en los atributos fecha y descripción, dado que no estaban presentes en el archivo original. Se ha cumplido con el objetivo definido en el avance anterior de procesar los datos en MatLab, por razones de complejidad no en R, extrayendo la información de los archivos XML y tabulando cada obra y sus atributos en una planilla de Excel que se cargó a R, para desde ahí realizar los 4 Challenges. Adjunto a este informe van los archivos de código Matlab, R y la planilla de la base de datos.

	1	2	3	4	5	6
	ID	Título	Artista	Descripción	Tipo	Fecha
1	'SK-A-4878'	'Een jonge ...	'schilder: E...	'Een jonge vro...	'schilderij'	'1644 - 1648'
2	'SK-A-4877'	'Portret van...	'schilder: ...	'Portret van Ja...	'schilderij'	'1857 - 1857'
3	'SK-A-4881'	'Portret van...	'schilder: ...	'Portret van ee...	'schilderij'	'1675 - 1685'
4	'RP-P-1992-...	'Keizer Karel...	'vermeld op...	'Keizer Karel V ...	'prent'	'1555 - 1555'
5	'RP-P-1992-...	'Gevangena...	'mogelijk pr...	'De gevangenn...	'prent'	'1555 - 1556'
6	'RP-P-1992-...	'Christus en...	'vermeld op...	'Op een plein i...	'prent'	'1553 - 1553'
7	'RP-P-1992-...	'Een vrouw ...	'prentmake...	'In een landsch...	'prent'	'1666 - 1706'
8	'RP-P-1992-...	'Vanitasstill...	'vermeld op...	'Allegorie op d...	'prent'	'1626 - 1626'
9	'BK-NM-48...	'Stortebeker...	'pottenbak...	'Stortebeker in ...	'stortebeker'	'1589 - 1589'
10	'BK-NM-10...	'Kruik met ...	' : anoniem'	'Kruik met oor ...	'kruik'	'ca. 1600 - c...
11	'BK-1991-30'	'Kom van k...	'pottenbak...	'Kom van kwar...	'kom'	'ca. 1200 - c...
12	'BK-NM-57'	'Schotel...	' : anoniem'	'Schotel af ...	'schotel'	'ca. 1700 - c...

	1	2	3	4	5	6
	ID	Título	Artista	Descripción	Tipo	Fecha
112029	'AK-RBK-19...	'Vaas besch...	'pottenbak...	'Vaas van pors...	'vaas'	'ca. 1680 - c...
112030	'AK-RBK-19...	'Vaas besch...	'pottenbak...	'Vaas van pors...	'vaas'	'ca. 1680 - c...
112031	'AK-RBK-19...	'Vaas besch...	'pottenbak...	'Vaas van pors...	'vaas'	'ca. 1680 - c...
112032	'AK-RBK-19...	'Vaas besch...	'pottenbak...	'Vaas van pors...	'vaas'	'ca. 1680 - c...
112033	'AK-RBK-19...	'Vaas besch...	'pottenbak...	'Vaas van pors...	'vaas'	'ca. 1680 - c...
112034	'AK-RBK-19...	'Vaas besch...	'pottenbak...	'Vaas van pors...	'vaas'	'ca. 1680 - c...
112035	'AK-RBK-14...	'Schotel bes...	'pottenbak...	'Schotel van p...	'schotel'	'ca. 1700 - c...
112036	'RP-P-08-8...	'Beschiedin...	'vermeld op...	'Het Nederlan...	'prent'	'1863 - 1863'
112037	'NG-NM-77...	'Inktpot me...	' : anoniem'	'Inktpot, recht...	'inktpot'	'ca. 1590 - c...
112038	'NG-NM-83...	'Hartsvange...	'zwaardveg...	'Hartsvanger ...	'hartsvanger'	'1774 - 1790'
112039	'NG-KOG-2...	'Zilveren st...	' : George'	'Vijftig zilveren...	'munt'	'1808 - 1808'

Imagen: Pantallazo de datos tabulados.