IEDB Custom Export Update

Presented by: Kelly Wheeler, Senior Software Engineer

What is the current export?

Currently .csv only with dual headers

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Epitope	Related O	Related O	Related
2	Epitope ID	Object Type	Description	Epitope M	Epitope Modification(s)	Starting Pi	Ending Po	Non-pept	Epitope Sy	Antigen Name	Antigen Accession	Parent Protein	Parent Protein Accession	Organism Name	Parent Organism	Parent Organism ID	Epitope Cl	Epitope Re	Object Ty	Descrip
3	58560	Linear peptide	SIINFEKL			258	265			ovalbumin	AAA48998.1	Gal d 2	P01012	Gallus gallus	Gallus gallus	9031				
4	114230	Linear peptide	SIINFEKL + SCM(K7)	K7	Side chain modification	258	265			ovalbumin	AAA48998.1	Gal d 2	P01012	Gallus gallus	Gallus gallus	9031	The epitope is modified with trinitroph			
5	1309579	Linear peptide	SIINFEKL + BIOT(L8)	L8	Biotin|BIOT	257	264			ovalbumin	0705172A	Gal d 2	P01012	Gallus gallus	Gallus gallus	9031				
6																				
7																				
8																				

```
1  Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Related Object,Related Object,Related Object,Related Object,Related Object,Related Object,Related
2  Epitope ID,Object Type,Description,Epitope Modified Residue(s),Epitope Modification(s),Starting Position,Ending Position,Non-peptidic epitope Accession,Epitope Synonyms,Antigen Name,Antigen Accession,Parent Protein,Parent Protein Acc
3  "58560","Linear peptide","SIINFEKL","","","258","265","","","ovalbumin","AAA48998.1","Gal d 2","P01012","Gallus gallus","Gallus gallus","9031","","","
```

Why change?

- One of the most requested change to the IEDB!
- Limited to .csv file format
- Fixed set of export fields
- Dual headers only
- Limited explanation of what each field means
- Limited information on the query performed

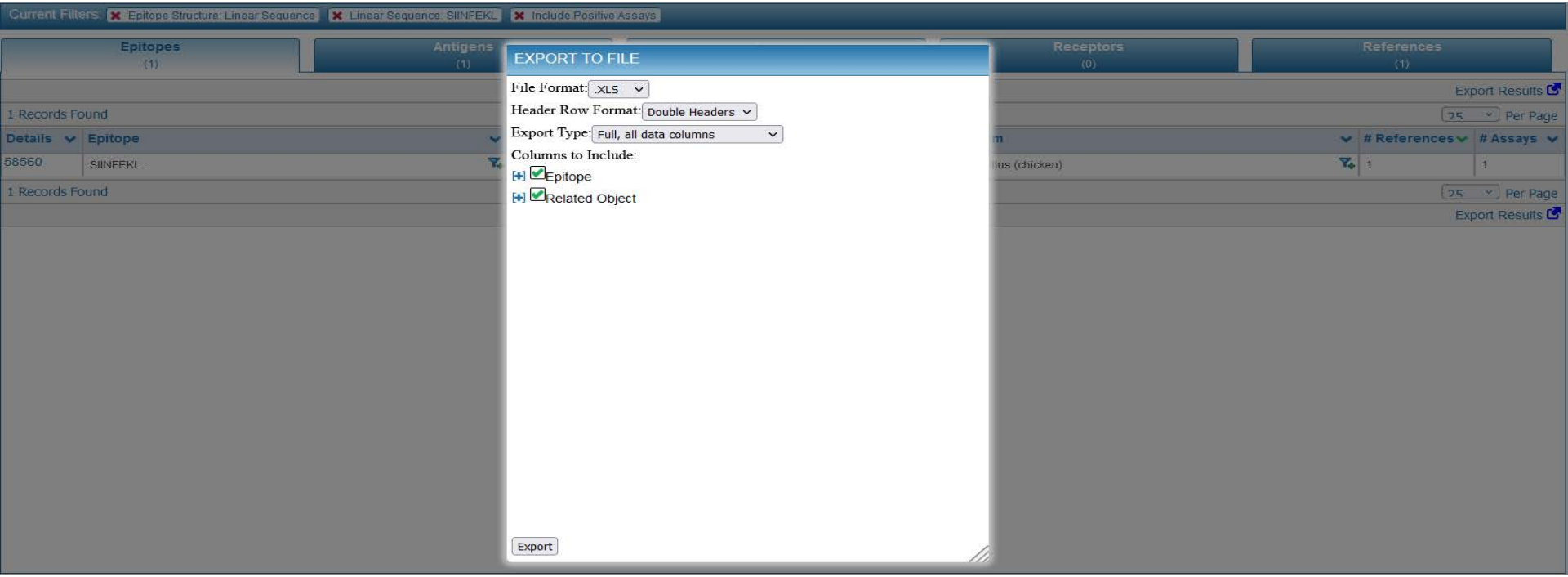
Introducing the new Custom Exports

Goals:

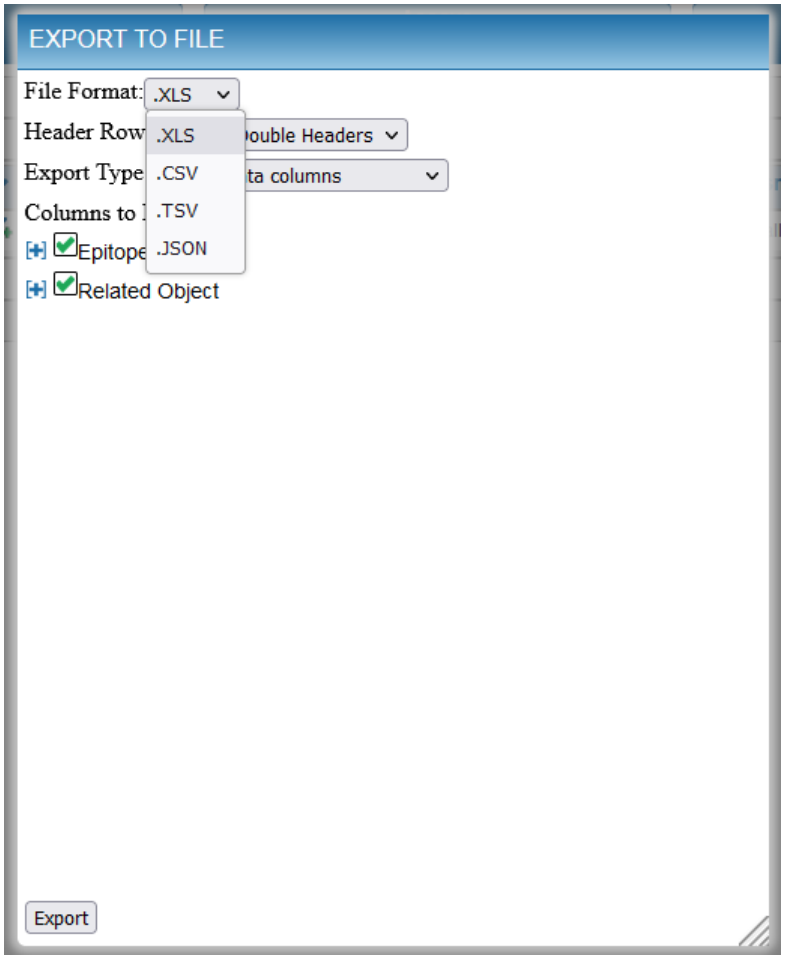
- Expand the usability of the exports
- Provide more information on each field
- Allow more customization to the user
- Provide more information linking the exported data to the query ran

* All shown functionality is currently in development and subject to change

New Custom Exports



Multiple File Formats



.XLS – Excel format – Most user friendly

Epitope														
Epitope ID	Epitope IRI	Object Type	Description	Epitope Modified Residue(s)	Epitope Modification(s)	Starting Position	Ending Position	Non-peptidic epitope ChEBI ID	Non-peptidic epitope IRI	Epitope Synonyms	Antigen Name	Antigen Accession	Antigen IRI	Parent Protein
58560	http://www.iedb.	Linear peptide	SIINFEKL			257	264			ovalbumin	ovalbumin	0705172A	http://www.	Gal d 2

.CSV – Comma Separated Values – Current format – Programmer friendly

```
Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,Epitope,"Related Object","Related Object","Related Object
"Epitope ID","Epitope IRI","Object Type",Description,"Epitope Modified Residue(s)","Epitope Modification(s)","Starting Position","Ending Position","Non-peptidic epitope ChEBI ID","Non-peptidic epitope IRI","Epitope Synonyms","Antigen
58560,http://www.iedb.org/epitope/58560,"Linear peptide",SIINFEKL,,,257,264,,,ovalbumin,ovalbumin,0705172A,http://www.ncbi.nlm.nih.gov/protein/0705172A,"Gal d 2",P01012,http://www.uniprot.org/uniprot/P01012,"Gallus gallus",http://pur
```

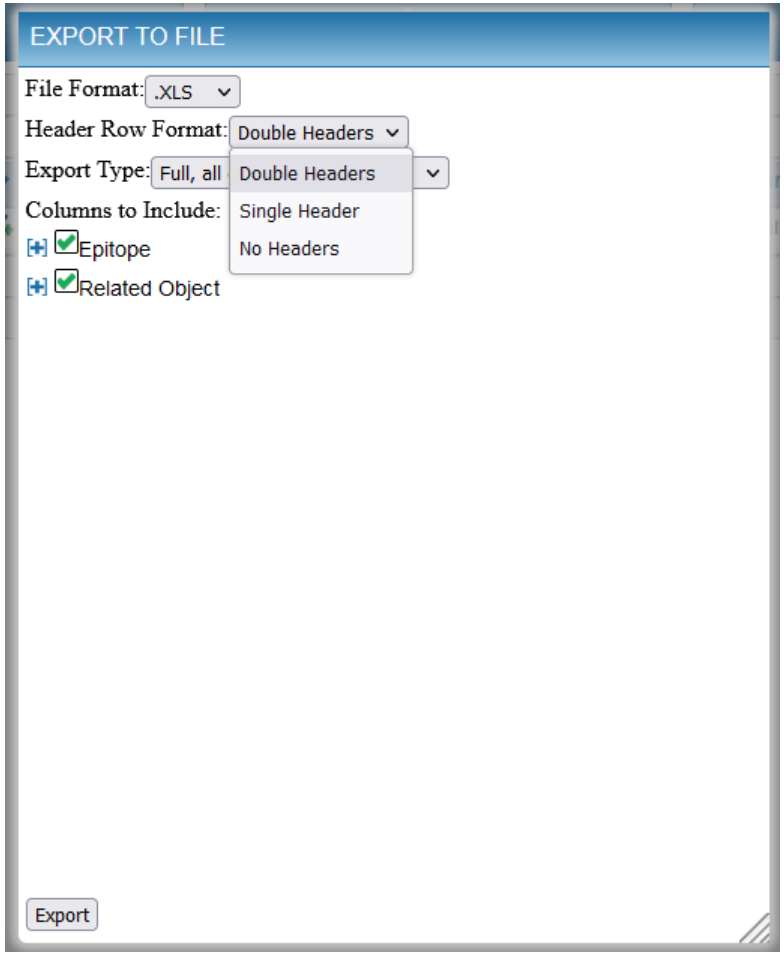
.TSV – Tab Separated Values – Similar to CSV but with tabs – Programmer friendly

```
Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope Epitope "Related Object"    "Related Object"    "Related
"Epitope ID"    "Epitope IRI"    "Object Type"    Description "Epitope Modified Residue(s)"    "Epitope Modification(s)"    "Starting Position" "Ending Position"    "Non-peptidic epitope ChEBI ID" "Non-peptidic epitope IRI"  "Epitope Sync
58560    http://www.iedb.org/epitope/58560    "Linear peptide"    SIINFEKL            257 264         ovalbumin    ovalbumin    0705172A    http://www.ncbi.nlm.nih.gov/protein/0705172A    "Gal d 2"    P01012    http://www.uniprot.org/uniprc
```

.JSON – JavaScript Object Notation – A programming friendly format used to transport data

```
{
    "Search Parameters":[
        "Epitope Structure: Linear Sequence",
        "Linear Sequence: SIINFEKL",
        "Include Positive Assays"
    ],
    "Export Date":"October 22, 2021",
    "Data":[
        {
            "Epitope - Epitope ID":"58560",
            "Epitope - Epitope IRI":"http:\/\/www.iedb.org\/epitope\/58560",
            "Epitope - Object Type":"Linear peptide",
            "Epitope - Description":"SIINFEKL",
            "Epitope - Epitope Modified Residue(s)":"",
            "Epitope - Epitope Modification(s)":"",
            "Epitope - Starting Position":"257",
            "Epitope - Ending Position":"264",
            "Epitope - Non-peptidic epitope ChEBI ID":"",
            "Epitope - Non-peptidic epitope IRI":"",
            "Epitope - Epitope Synonyms":"ovalbumin",
            "Epitope - Antigen Name":"ovalbumin",
            "Epitope - Antigen Accession":"0705172A",
```

Multiple Header Formats



Double Headers – Current format with a section and field header row

	A	B	C	D	E	F	G	H	
1	Epitope								
2	Epitope ID	Epitope IRI	Object Type	Description	Epitope Modified Residue(s)	Epitope Modification(s)	Starting Position	Ending Position	Non-peptidic
3	58560	http://www.	Linear peptide	SIINFEKL			257	264	

Single Header – Single header row with section and field merged

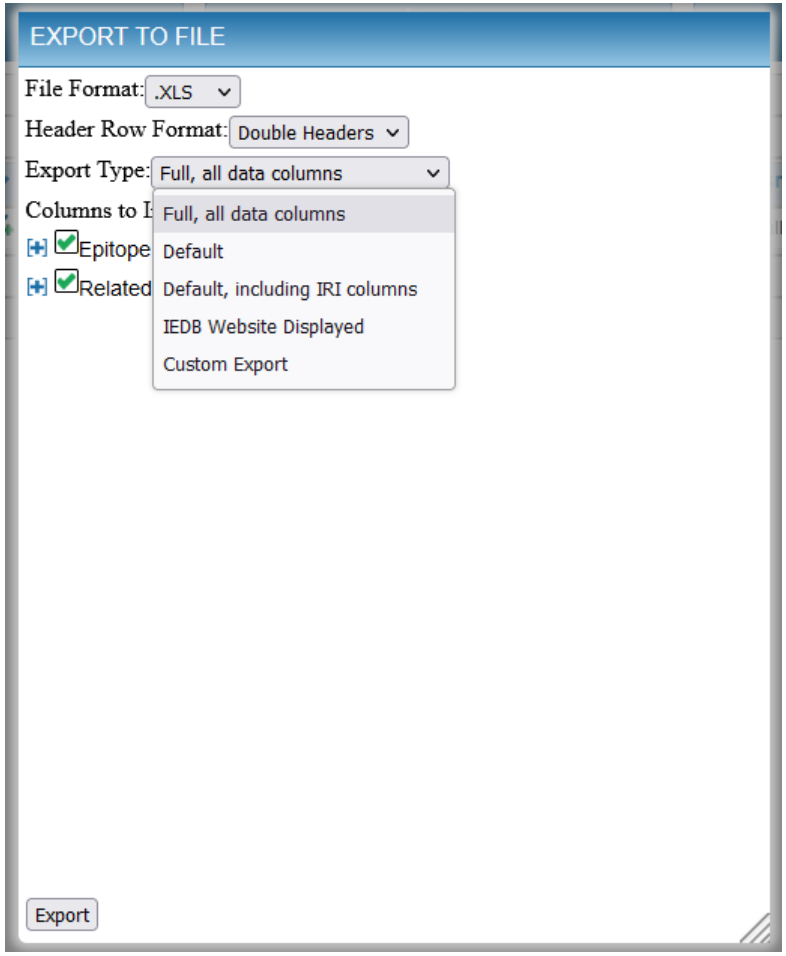
	A	B	C	D	E	F	G	H	
1	Epitope								
2	Epitope ID	Epitope IRI	Object Type	Description	Epitope Modified Residue(s)	Epitope Modification(s)	Starting Position	Ending Position	Non-peptidic
3	58560	http://www.	Linear peptide	SIINFEKL			257	264	

Double Headers – Current Format with a section and field header row

	A	B	C	D	E	F
1	Epitope - Epitope ID	Epitope - Epitope IRI	Epitope - Object Type	Epitope - Description	Epitope - Epitope Modified Residue(s)	Epitope - Epitope Modification(
2	58560	http://www.iedb.org/	Linear peptide	SIINFEKL		
3						

No Headers – Simple data centric format with no headers

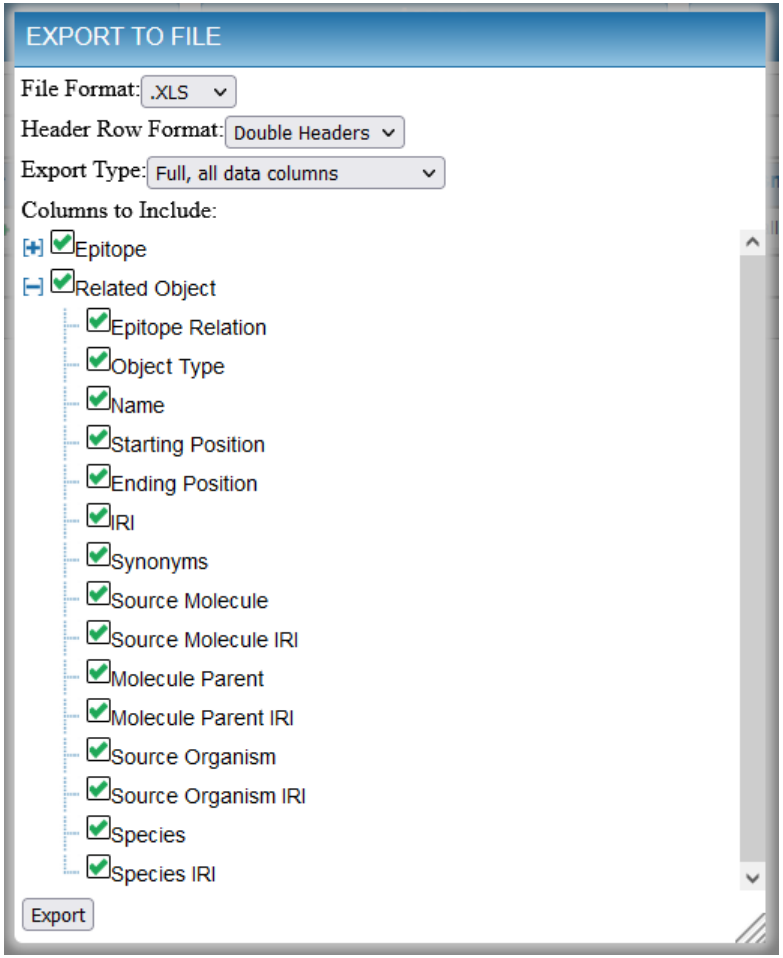
Custom Field Selection and Export Types



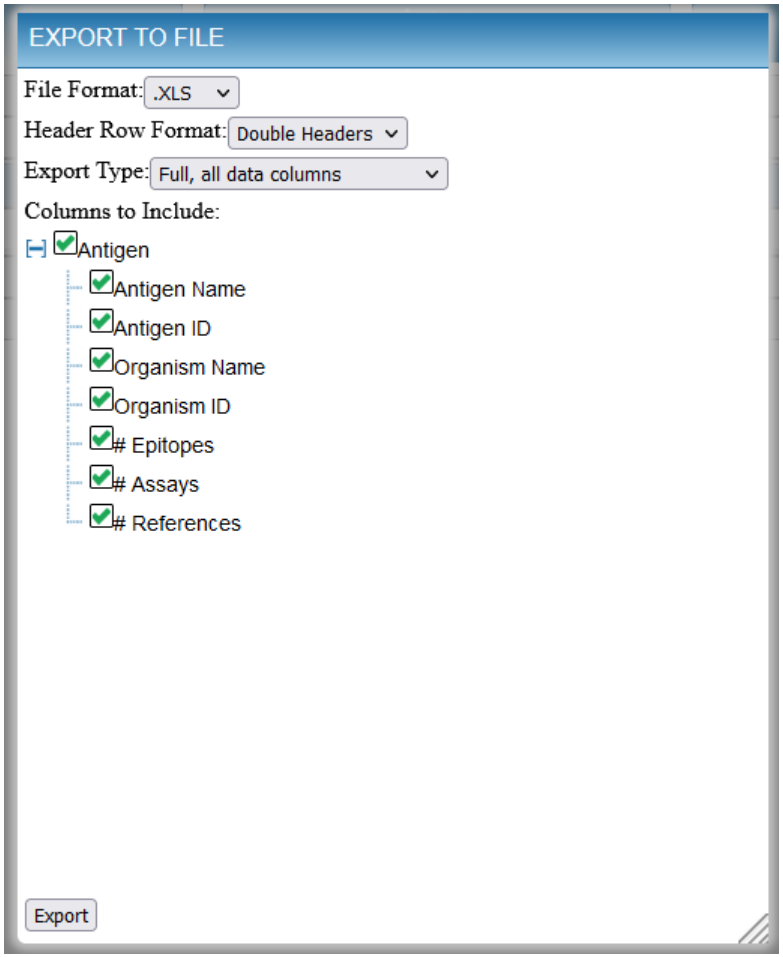
- Export types are preselected subsets of fields to be exported. These are currently examples only and will be further refined upon release.
- Custom field section shown in a similar structure as the finder trees.
- Fields shown under section headers that expand to show the individual field names that can also be selected separately.
- Allows for extremely fine grain customization of exports to fit each users needs.

Field Selection Examples

Epitope

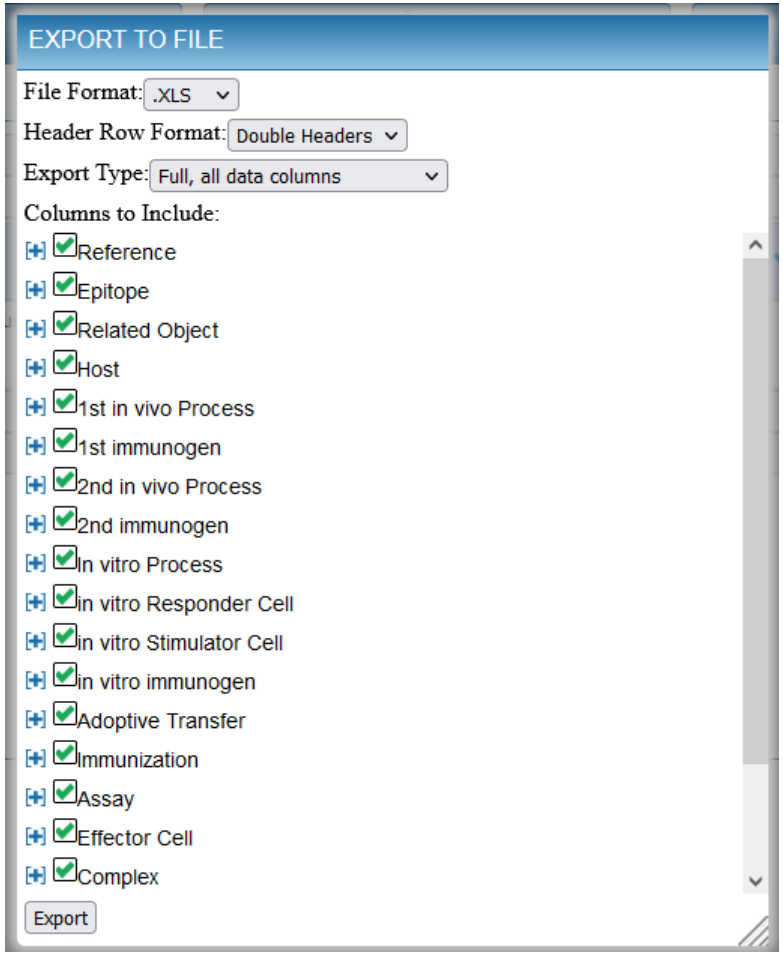


Antigen

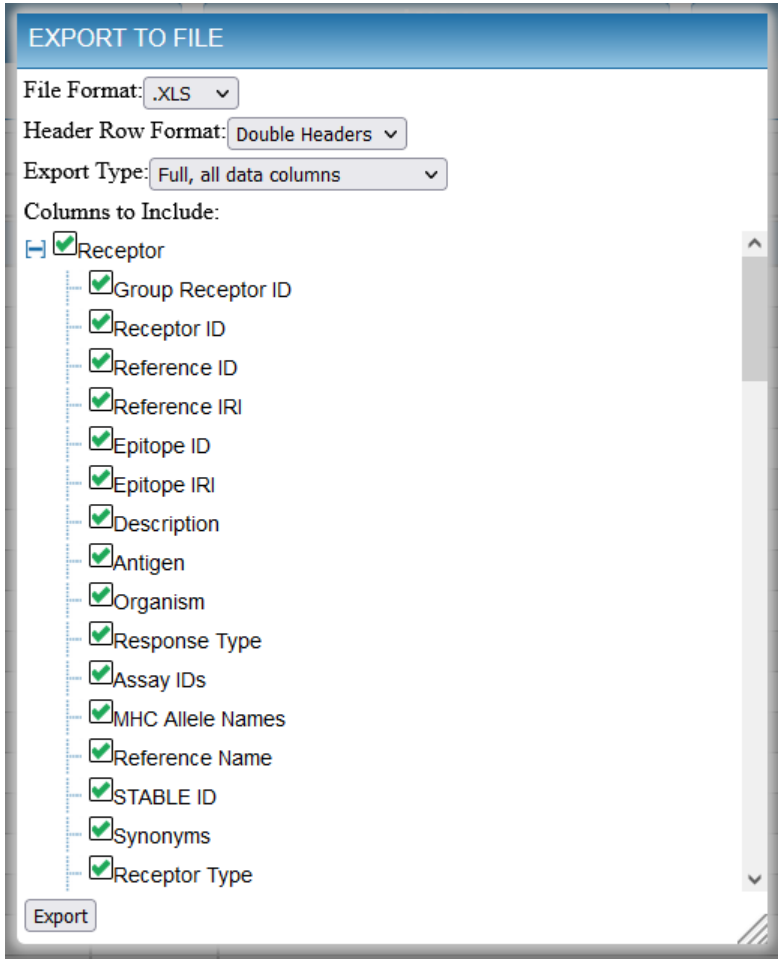


Field Selection Examples

Assay

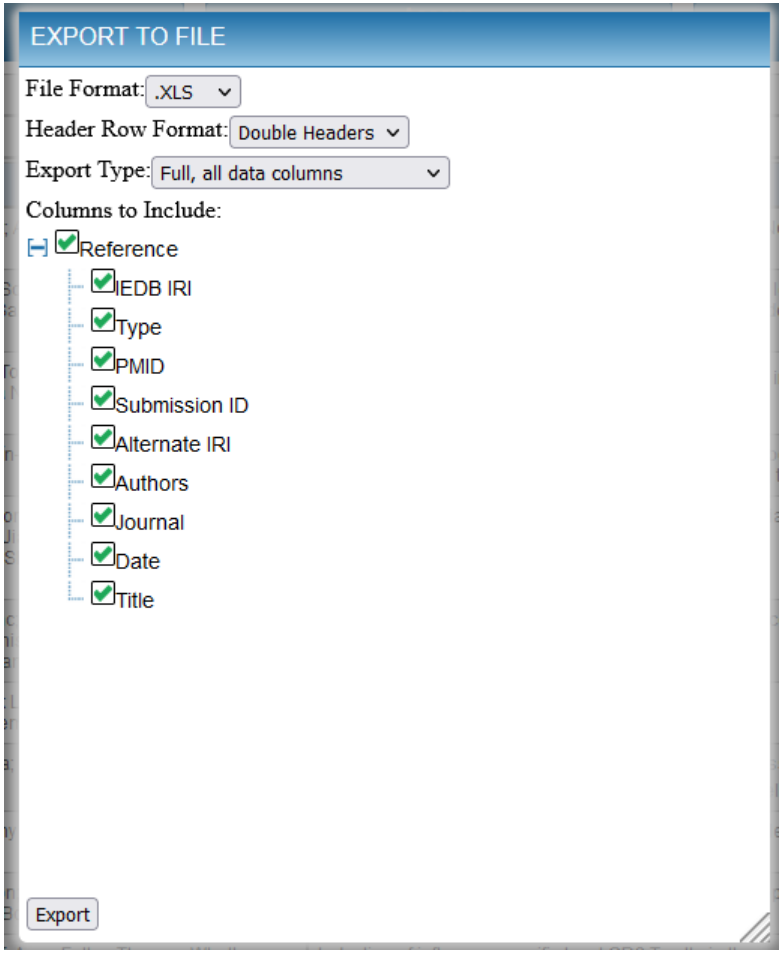


Receptor



Field Selection Examples

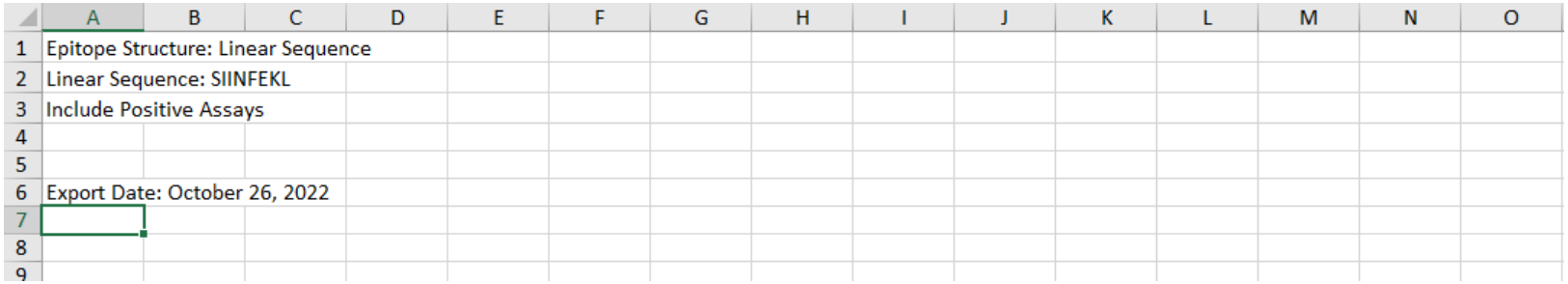
Reference



Additional Improvements

- Inclusion of query parameters in the Excel export via a new sheet
- Inclusion of field descriptions in the Excel export via a new sheet
- Re-evaluation of available fields and field names
- Addition of the option to export the fields as shown in the tabs on the website

Export Parameters Example



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Epitope Structure: Linear Sequence														
2	Linear Sequence: SIINFEKL														
3	Include Positive Assays														
4															
5															
6	Export Date: October 26, 2022														
7															
8															
9															

- Includes the full list of the parameters used on the IEDB website to generate this export
- Also will include the date the export was ran for future use.
- Available in the Excel and JSON formats.

Field Help Tab Example

	A	B	C	D	E	F	G	
1	Section							Epito
2	Header	IEDB IRI	Object Type	Name	Modified Residue(s)	Modifications	Starting Position	Ending Pos
3	Example	http://www.iedb.org/epitope/1056	Linear peptide	AELLVALENQHTIDL			441	455
4	Example	http://www.iedb.org/epitope/2136	Linear peptide	AIYHTENAYVSVVSSHYNR			207	225
5	Other possible values	http://www.iedb.org/epitope/570	Linear peptide	ACKRGPGSGFFSRLN	R4	Citrullination|CITR	138	152
6	Description of Field	The url for this epitope on the IEDB site.	The object type of the epitope, which may include Linear and	A short description of the epitope, including its sequence or	The amino acid and its position within	The type of post-transla	The starting position of	The ending
7	Help Topic Link		http://curationwiki.iedb.org/wiki/index.php/Data_Field_Des	http://curationwiki.iedb.org/wiki/index.php/Data_Field_Des	http://curationwiki.iedb.org/wiki/index	http://curationwiki.iedb.	http://curationwiki.iedb.	http://cura
8								
9								
10								

- A new tab in the Excel export that offers information on each field selected and exported.
- Each field has the section and header name matching the header format selected.
- Below this are 2 pre-selected examples chosen by our team to demonstrate what is in the IEDB.
- After this is another possible value to show what that field could include in case our above examples are blank or we want to show other possible values.
- Next is a brief description of what that field represents.
- Finally there is a link to an IEDB help article for that field if available.