Copy / Paste out of browser after ArtNet search:

In original search window, or Print browser window, Select all, then copy and paste into TextWrangler (or other text editor with Regular Expression find / replace)

PC Text Editor (e.g. Notepad++) Cleanup (PC lines end with \r\n)

Blank lines:

```
^[ \t]*\r\n
```
replace with nothing

Extra lines and artist name:

```
Full details\t\r\n(\d+)\t\r\n \t (.*)\r\n
```
replace with
```
Artist\t\2\t\1\r\n
```

Continued lines (don't contain tabs):

```
\r\n([^\t]*)\r\n
```
replace with (space, not really underscore)
```
_\1\r\n
```

Mac TextWrangler Cleanup (Mac lines end with \r)

Blank lines:

```
^[ \t]*\r
```
replace with nothing

Extra lines and artist name:

```
Full details\t\r(\d+)\t\r \t (.*)\r
```
replace with
```
Artist\t\2\t\1\r
```

Continued lines (don't contain tabs):

```
\r([^\t]*)\r
```
replace with (space, not really underscore)
```
_\1\r
```

Google Refine Processing

Convert to one column per field format

Create new project and load text file into Google Refine as TSV file
Uncheck "use 1 column as headers" before creation
On column 3, Edit Cells -> Fill Down
Transpose -> Columnize on key/value columns… with Column 1 as key and Column 2 as value
Export as Excel file or TSV

Create a new column with "sold for" number amount by itself

On "Sold For", click the triangle menu: Edit Column -> Add column based on this column…
```
title: price_USD
expression: value.match(/(^|.*[ ])([0-9,]+) USD.*/)[1]
```

Brief explanation: The pipe | character means "or", so it looks for either the beginning of the line ^ or other characters .* followed by either a space or an opening parenthesis, then digits or a comma followed by USD. Since the first piece is in parentheses, it captures that (I could put a ?: inside the parentheses to make them "non-capturing"), so the dollar amount is the 2nd piece I'm capturing, so I need to use the [1] to get the 2nd list element.

Create a new column with "auction date" by itself

On the "Sale of" column, click the triangle menu: Edit Column -> Add column based on this column…
```
title: sale_date
expression: value.match(/(.*): [A-Za-z]+, (.*)\[(.*)/)[1]
```

Export as TSV for better Excel interpretation of numbers and dates

While importing into Google Refine, in TSV options, after unchecking "Parse next 1 line(s) as column headers", also uncheck "Quotation marks are used to enclose cells containing column separators". Problems arise if unmatched double quotes are part of data like Description -- treats rest of data as single cell.

Auction house
value.match(/(.*?):.*/)[0]

Estimate range USD (from Estimate column -- creating this intermediate column to make it easier to extract low and high estimate values below)
value.match(/(^|.*[ ])([0-9,]+ - [0-9,]+) USD.*/)[1]

Estimate low USD (from estimate range)
value.match(/([0-9,]+) - ([0-9,]+)/)[0]

Estimate high USD (from estimate range)
value.match(/([0-9,]+) - ([0-9,]+)/)[1]

Edition (2nd number after slash. For first number replace [1] with [0])
value.match(/(\d+)\/(\d+)/)[1]

Beware of "Style of" artist lines -- different format that typical artist lines

Google Refine allows you to save a series of operations and reapply them to a new data set
https://code.google.com/p/google-refine/wiki/History