

## processed 資料夾：過程檔案

LibreOffice 7.0.6.2 (x64)

<https://www.libreoffice.org/>

Notepad++ 7.9.5 (64-bit)

<https://notepad-plus-plus.org/>

Python 3.9.6

<https://www.python.org/>

輸入檔案	處理過程	輸出檔案
02.csv	n02()	02.xlsx
03.pdf	n03()	03.xlsx
04.pdf	n04()	04.xlsx
05.pdf	n05()	05.xlsx
82.doc	以 LibreOffice Writer 開啟複製 在 LibreOffice Calc 以 Unformatted Text · Separated by: Space, Other(、)貼上	82.xlsx
08.pdf 08_2.pdf	n08()  參考以下連結，cid 尚無普遍有效解決方法，因此捨去 ChemicalChnName 欄位 <a href="https://stackoverflow.com/questions/66656067/replace-cidnumber-with-&lt;br/&gt;chars-using-python-when-extracting-text-from-pdf-fil">https://stackoverflow.com/questions/66656067/replace-cidnumber-with- chars-using-python-when-extracting-text-from-pdf-fil</a>	08.xlsx
09.pdf 09_2.pdf	n09()	09.xlsx
10.pdf 10_2.pdf 10_3.pdf	n10()	10.xlsx
11.pdf 11_2.pdf	n11()	11.xlsx
12.odt	以 LibreOffice Writer 開啟複製表格 到 LibreOffice Calc 以 Rich text formatting (RTF)貼上	12_1.xlsx
12_1.xlsx	n12()	12.xlsx
13.pdf	n13()	13.xlsx
14.xlsx	n14()	14.xlsx
15.xlsx	n15()	15.xlsx
16.xlsx	n16()	16.xlsx
17.xlsx	n17()	17.xlsx
19.pdf	n19()	19.xlsx
20.pdf	n20()	20.xlsx
21.pdf	n21()	21.xlsx
22.doc	以 LibreOffice Writer 開啟選擇相關資料複製 到 LibreOffice Calc 貼上自行整理	22.xlsx
24.odt	以 LibreOffice Writer 開啟選擇相關資料複製 到 LibreOffice Calc 貼上自行整理	24.xlsx

27.xlsx	n27()	27.xlsx
28.xlsx	n28()	28.xlsx
29.xlsx	n29()	29.xlsx
30.xlsx	n30()	30.xlsx
31.pdf	n31()	31.xlsx
32.pdf	n32()	32.xlsx
33.pdf	pdf 讀不到字  以下列連結複製表格 Ministry of the Environment, Government of Japan Strategic Programs on Environmental Endocrine Disruptors '98 (SPEED '98) <a href="http://www.env.go.jp/en/chemi/ed/speed98/sp98t3.html">http://www.env.go.jp/en/chemi/ed/speed98/sp98t3.html</a> Table-3: Chemicals Suspected of Having Endocrine Disrupting Effects 在 LibreOffice Calc 以 Simple HTML 貼上 將 Substances 欄位資料剪下貼到 Notepad++ 將 Notepad++ 資料剪下貼到 LibreOffice Calc Seperated by: Tab, other(.), trim spaces 對照 pdf 調整 11 保留 Simazine 13 保留 Carbaryl 32 保留 Camphechlor 44 更正 66 刪除 67 刪除	33_1.xlsx
33_1.xlsx	n33()	33.xlsx
34.xlsx	n34()	34.xlsx
35	複製網頁內容 在 LibreOffice Calc 貼上自行整理 CASNo 來自清單中連結 2021-09-28 執行	35.xlsx
36.xlsx	參照表 Notes 在 LibreOffice Calc 自行整理 CASNo 對應表	36_1.xlsx
36.xlsx 36_1.xlsx	36()	36.xlsx
37	複製網頁表格 在 LibreOffice 以 Unformatted Text , Seperated by: Tab, Trim spaces , Column type: Text 貼上 調整 Cobalt(II) oxide 位置	37_1.xlsx
37_1.xlsx	n37()	37.xlsx
38	複製網頁表格 在 LibreOffice 以 Unformatted Text , Seperated by: Tab, Trim spaces , Column type: Text 貼上	38_1.xlsx
38_1.xlsx	n38() 有日文名稱但同時缺乏 ChemicalEngName 及 CASNo 者在以上處理中剔除	38.xlsx
39	複製網頁表格 在 LibreOffice 以 Unformatted Text , Seperated by: Tab, Trim spaces , Column type: Text 貼上	39_1.xlsx

39_1.xlsx	n39()	39.xlsx
40_43.csv	n40_43()	40.xlsx 41.xlsx 42.xlsx 43.xlsx
44_45.xlsx	n44_45()	44.xlsx 45.xlsx
46.xlsx	n46()	46.xlsx
48.doc	以 LibreOffice Writer 開啟複製表格 在 LibreOffice Calc 以 HTML · language: automatic 貼上 註：以 RTF 貼上 CASNo 有些會是日期	48_1.xlsx
48_1.xlsx	n48()	48.xlsx
49	複製網頁表格 在 LibreOffice 以 Unformatted Text · Seperated by: Tab, Trim spaces · Column type: Text 貼上 註：以 Simple HTML 貼上 CASNo 有些會是日期	49_1.xlsx
49_1.xlsx	n49()	49.xlsx
83.pdf	n83()	83.xlsx
50.pdf	n50()	50.xlsx
53_55.doc	以 LibreOffice Writer 開啟複製 在 LibreOffice Calc 貼上自行整理	53.xlsx 54.xlsx 55.xlsx
56_57.pdf	n56_57()	56.xlsx 57.xlsx
58	複製網頁表格 在 LibreOffice Calc 以 Simple HTML 貼上 全選表格，formatting 為 text	58_1.xlsx
58_1.xlsx	n58()	58.xlsx
59	複製網頁表格 在 LibreOffice Calc 以 Simple HTML 貼上 全選表格，formatting 為 text	59_1.xlsx
59_1.xlsx	n59()	59.xlsx
60	複製網頁表格 在 LibreOffice Calc 以 Simple HTML 貼上 全選表格，formatting 為 text	60_1.xlsx
60_1.xlsx	n60()	60.xlsx
61	複製網頁表格 在 LibreOffice Calc 以 Simple HTML 貼上	61_1.xlsx
61_1.xlsx	n61()	61.xlsx
62	複製網頁表格 在 LibreOffice Calc 以 Simple HTML 貼上	62_1.xlsx
62_1.xlsx	n62()	62.xlsx
63	複製網頁表格 在 LibreOffice 以 Unformatted Text · Seperated by: Tab, Trim spaces · Column type: Text 貼上	63_1.xlsx
63_1.xlsx	n63()	63.xlsx

64	複製網頁表格 在 LibreOffice Calc 以 Unformatted Text , Seperated by: Tab, Trim spaces , Column type: Text 貼上	64_1.xlsx
64_1.xlsx	n64()	64.xlsx
65_67.odt	以 LibreOffice Writer 開啟複製 Substance 清單 在 LibreOffice Calc 以 Unformatted Text , Trim spaces 貼上	65_67.xlsx
65_67.xlsx	n65_67()	65.xlsx 66.xlsx 67.xlsx
68	複製網頁表格 在 LibreOffice Calc 以 Unformatted Text , Seperated by: Tab, Trim spaces , Column type: Text 貼上	68_1.xlsx
68_1.xlsx 68_2.pdf	n68()	68.xlsx
69_2	在 LibreOffice Calc 自行整理	69_2.xlsx
69.csv 69_2.xlsx	n69()	69.xlsx
72.pdf	n72()	72.xlsx
74	複製網頁表格 在 LibreOffice Calc 以 Unformatted Text , Seperated by: Tab, Comma, Other(.)貼上	74.xlsx
76.doc	以 LibreOffice Writer 開啟複製表格 在 LibreOffice Calc 以 HTML , language: automatic 貼上	76_1.xlsx
76_1.xlsx	n76()	76.xlsx
77	複製網頁表格 在 LibreOffice Calc 以 Simple HTML 貼上	77_1.xlsx
77_1.xlsx	n77()	77.xlsx
78	複製網頁清單 在 LibreOffice Calc 以 Unformatted Text , Trim spaces 貼上	78_1.xlsx
78_1.xlsx	n78()	78.xlsx
79.pdf	n79()	79.xlsx
80.doc 80_2.doc	以 LibreOffice Writer 開啟複製表格 在 LibreOffice Calc 以 HTML , language: automatic 貼上	80_1.xlsx
80_1.xlsx	n80()	80.xlsx
清單.xlsx 所有 processed 資料夾中 檔名不含_ 的檔案	Merged()	Merged.xlsx