

CODING FOR HUMANITIES

IMPORTANT DATA FORMATS AND DATA CLEANING, INTRODUCTION TO APIS, SHORT INTRODUCTION TO R

Dr. Thomas Köntges,
Digital Humanities, University of Leipzig

26 October 2015

SOME ACRONYMS

OCR, NER, NLP

OCR Optical Character Recognition

NER Named Entity Recognition

NLP Natural Language Processing

DIGITISATION

WHY DIGITISE?

Reasons given by the British Library

- Open up access to content in the British Library's collection for researchers
- Add value to, and open up previously unimagined areas for research
- Support innovative methods of research
- Facilitate the interpretation of our content by others for new audiences
- Preserve unique, rare and fragile heritage items by digital reproduction and protect vulnerable documents
- Reveal illegible and hidden text or images and permit non-intrusive testing of materials

FORMATS

TIFF Tagged Image File Format by Adobe

TXT #iloveplaintxt

CSV Comma Separated Value

PDF Portable Document Format (Open Standard)

TSV Tab Separated Value

XML Extensible Markup Language

JSON JavaScript Object Notation

FORMATS

TXT

1834 November 1: to Coates
Sydney, New South Wales, Novr. 1, 1834.
To Dandeson Coates Esqr.
Dr. Sir,
By the blessing of God we are safely arrived at this "ha"
I premise, that I need not enter into a detail of Incide
send you a full account of whatsoever he deems worthy you
expected, thanks to the Almighty, the Society, and Capt.
I trust, Sir, that the Passengers who accompanied us, and
in her.-I am, I trust, thankful for His mercies, and I do
In a letter which I received from the C.M.H., bearing dat

CSV

Collector 1,Collector 2,Date Collected,Identified By (1),Identified By (2),Identified By (3),Identified (4),Identified (5),Identified Date (1),Identified Date (2),Identified Date (3),Taxon,Location 1,IRN,Prefix Souce

Date	Date (year)	Addressee	Location
May 11, 1871	1878	D. Black	Napier
May 11, 1871	1878	J.McCulloch	Napier
May 21, 1871	1878	Luff	Napier
Jun 12, 1878	1878	Luff	Napier
Jul 9, 1878	1878	Haast	Napier
Jul 18, 1878	1878	Luff	Napier
Aug 20, 1878	1878	Balfour	Napier
Jan 1, 1878	1878	Balfour	Napier
Undated	Undated	Harding	Not named
Nov 23, 1878	1878	Mantell	Not named

<arg>
<p><hi rend="i"><name type="work" key="http://wtap.vuw.ac.nz/eats/entity/331/">Poems o
3997/">G. Robertson & Co.</name>, <name type="place" key="http://wtap.vuw.ac.n
5250/">Sydney</name>, &c.</p>
</arg>
<p><hi rend="sc">A reproach</hi> has been removed from Australian literature by the pul
4141/">Henry Clarence Kendall</name>. A generation hence, his works will probably b
day, and <name type="person" key="http://wtap.vuw.ac.nz/eats/entity/4141/">Kendall

name type="place" key="http://wtap.vuw.ac.nz/eats/entity/4201/">Australia</name> ha
contemporaries the poet and his genius were alike matters of indifference. The tast

1994 TEI



JSON

```
X Developers - API docs v3 - Search Rec... API request - thomas.koentges@gmail.com api.digitalnz.org/v3/records.json?api_... api.digitalnz.org/v3/records.json?api... Create a Scrolling Effect in PowerPoi... +  
Developers - API docs v3 - Search Records  
{"id":22702847,"updated_at":"2015-06-18T18:56:26.709+12:00","created_at":"2012-04-21T07:23:26.000+12:00","title":"Scott, Thomas, 1947- : [Cnristcnurcn earthquake] 25 February 2011","description":"The cartoon shows a digger dredging through the rubble and digging up a red heart representing 'hope' (Tom Scott doesn't do colour so this is significant). A rescuer nearby yells 'Careful! It's still beating'. Context - on 22 February 2011 a 6.3 magnitude earthquake struck in Christchurch which has probably killed more than 200 people (at this point the number is still not known) and caused much more severe damage. There were many people trapped in collapsed buildings and it was apparent in only two or three days that in most cases they could not have survived but of course people still held out impossible hope.\nQuantity: 1 digital cartoon(s).\nPhysical Description: Image file - Jpeg","content_partner":["Alexander Turnbull Library"],"category":["Images"],"creator":["Not specified"],"dc_type":["Item"],"dnz_type":"Artwork","date":["2011-01-01 12:00:00 UTC"],"source_url":"http://api.digitalnz.org/records/22702847/source","collection":["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "TAPUHI", "Drawings and Prints Collection", "New Zealand Cartoon Archive", "CEISMIC"], "alternate_title":[],"additional_description":[],"display_content_partner":"Alexander Turnbull Library","collection_title":["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "TAPUHI", "Drawings and Prints Collection", "New Zealand Cartoon Archive", "CEISMIC"], "display_collection":"TAPUHI","primary_collection":["TAPUHI"], "contributing_partner":[],"contributor":[],"copyright":["All rights reserved"], "citation":[]}, "credit_creator":null, "language":["en"], "provenance":null, "publisher":[], "rights":"Please check copyright", "usage":["All rights reserved"], "source":[], "tag":[]}, "CEISMIC"], "thesis_level":null, "holding":[], "library_collection":["Drawings and Prints Collection", "New Zealand Cartoon Archive"], "shelf_location": "DCDL-0017168", "eprints_type":[]}, "text":null, "fulltext":null, "format":["Digital images", "Cartoons (Commentary)", "1 digital cartoon(s)", "Single art work", "Image file - Jpeg"], "dc_identifier": ["ndha:IE3303922", "tap:1410005", "urn:nbn:nz:wtu:DCDL-0017168", "DCDL-0017168"], "display_date": "2011", "published_date":[], "syndication_date": "2014-08-04T15:29:35.905+12:00", "landing_url": "http://natlib.govt.nz/records/22702847", "large_thumbnail_url": "http://ndhadeliver.natlib.govt.nz/NLNZStreamGate/get?dps_pid=IE3303922", "rights_url": []}, "thumbnail_url": "http://ndhadeliver.natlib.govt.nz/delivery/DeliveryManagerServlet?dps_pid=IE3303922\u0026dps_func=thumbnail", "origin_url": "http://tapuhi.natlib.govt.nz/cgi-bin/spydus/NAV/GLOBAL/OPHDR/1/1410005", "metadata_url": null, "object_url": null, "has_version":[], "license":null, "relation": ["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "DC-Group-0025"], "tap:1024981], "is_part_of": ["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "DC-Group-0025", "tap:1024981], "is_replaced_by":null, "replaces":null, "is_required_by":null, "is_version_of":null, "table_of_contents":null, "is_commercial_use":null, "atl_free_download":null, "atl_physical_viewable":[]}, "subject": ["Hope", "Christchurch Earthquake, N.Z., 2011", "Earthquakes", "New Zealand", "Canterbury Region", "Search and rescue operations", "Christchurch City"], "coverage":[], "attachments": [{"_id": "5556b31c646e7a5bbd396500", "aspect_ratio":null, "date":null, "dc_identifier": "IE3303922", "dc_type":null, "description":null, "display_date":null, "file_size":null, "file_type":null, "large_thumbnail_url": "http://ndhadeliver.natlib.govt.nz/delivery/DeliveryManagerServlet?dps_pid=IE3303922", "name": "25feb11.jpg", "ndha_rights":100, "thumbnail_url": "http://ndhadeliver.natlib.govt.nz/delivery/DeliveryManagerServlet?dps_pid=IE3303922\u0026dps_func=thumbnail", "title":null, "url":null}], "authorities": [{"_id": "55826b9a646e7a095201d628", "authority_id": 1137550, "name": "subject_authority", "role":null, "text": "Hope"}, {"_id": "55826b9a646e7a095201d629", "authority_id": 1411976, "name": "subject_authority", "role":null, "text": "Christchurch Earthquake, N.Z., 2011"}, {"_id": "55826b9a646e7a095201d62a", "authority_id": 144081, "name": "subject_authority", "role":null, "text": "Earthquakes - New Zealand - Canterbury Region"}, {"_id": "55826b9a646e7a095201d62b", "authority_id": 1601, "name": "subject_authority", "role":null, "text": "Search and rescue operations"}, {"_id": "55826b9a646e7a095201d62c", "authority_id": 65235, "name": "place_authority", "role":null, "text": "Christchurch City"}, {"_id": "55826b9a646e7a095201d62d", "authority_id": 942610, "name": "recordtype_authority", "role":null, "text": "Digital images"}, {"_id": "55826b9a646e7a095201d62e", "authority_id": 77590, "name": "recordtype_authority", "role":null, "text": "Cartoons (Commentary)"}, {"_id": "55826b9a646e7a095201d62f", "authority_id": 1024981, "name": "collection_parent", "role":null, "text": "Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post"]}, {"_id": "55826b9a646e7a095201d630", "authority_id": 1024981, "name": "collection_root", "role":null, "text": "Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post"]}, {"_id": "55826b9a646e7a095201d631", "authority_id": 864885, "name": "broad_related_authority", "role":null, "text": "Earthquakes - New Zealand"}, {"_id": "55826b9a646e7a095201d632", "authority_id": 266587, "name": "broad_related_authority", "role":null, "text": "Rescue work"}, {"_id": "55826b9a646e7a095201d633", "authority_id": 147273, "name": "broad_related_authority", "role":null, "text": "Civil defence"}, {"_id": "55826b9a646e7a095201d634", "authority_id": 1127936, "name": "broad_related_authority", "role":null, "text": "Public safety"}, {"_id": "55826b9a646e7a095201d635", "authority_id": 1226690, "name": "broad_related_authority", "role":null, "text": "Human services"}]
```

GOLD-RUSH: MINING

Text-Mining vs. **Data-Mining**

VISUALISATIONS

VISUALISATIONS

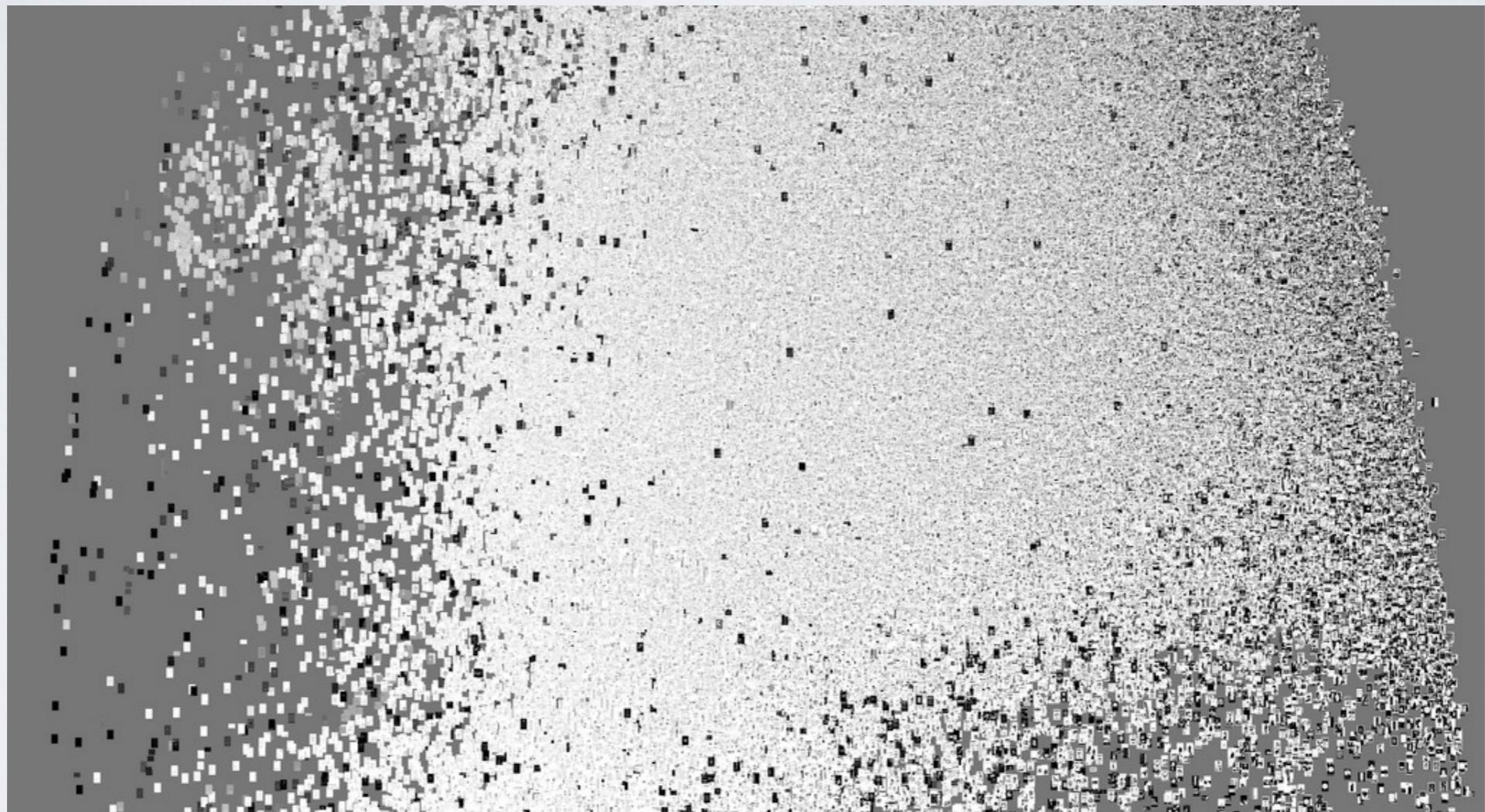
- history, visualisations, ideas, wisdom
- ἴστορία, ὕδειν, videre, weisheit
- sapientia, sapere aude (Horace, Kant) -> dare to taste
- we are using telescopes and microscopes to create knowledge
- macroscopes

MACROSCOPES

While microscopes help us to research objects that are too small for our eyes and telescopes things that are too far away, **macroscopes** will help us to **view and research things that are too big or complex** to comprehend.

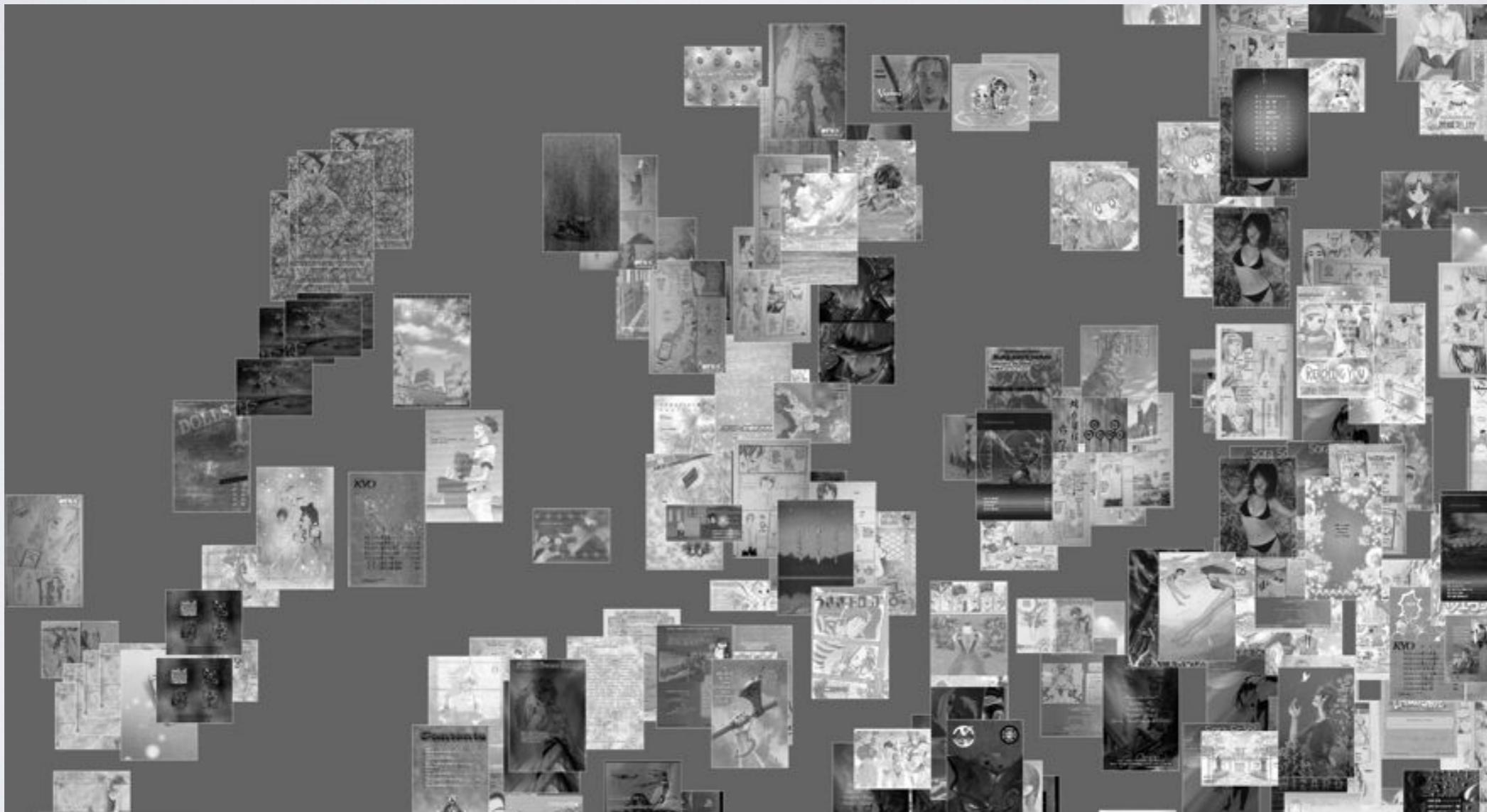
EXAMPLE

A Million Mangas: UCSD



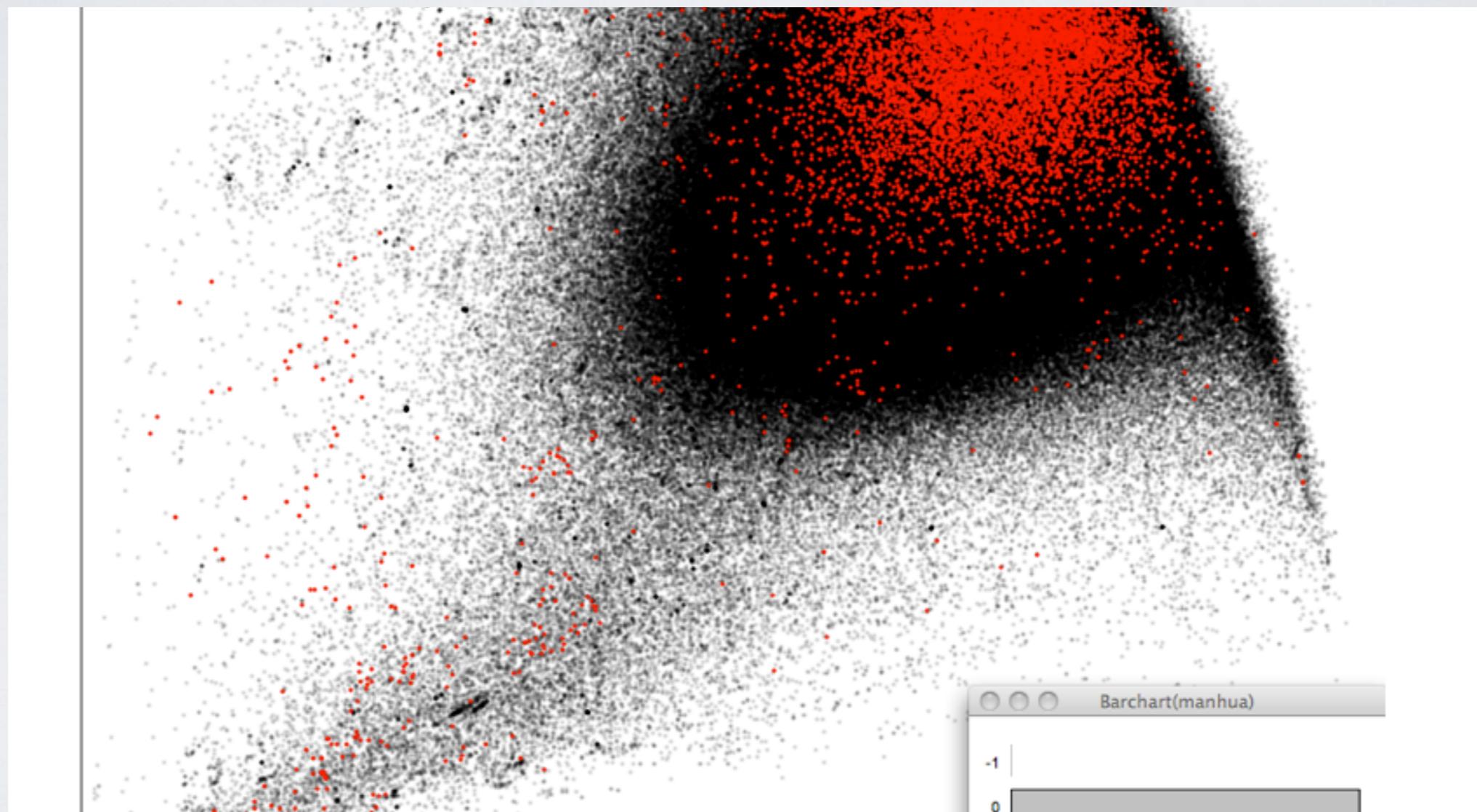
EXAMPLE

A Million Mangas: UCSD



EXAMPLE

A Million Mangas: UCSD



DIGITAL HUMANITIES RESEARCH

DIGITAL HUMANITIES RESEARCH

- Research question
- **Data**
- **Data cleaning**
- Analysis, Research question
- Visualisation (internal)
- Analysis, Research question
- Analysis, Research outcomes
- Visualisation (external)
- Repeat

DATA-CLEANING

TOOLS FOR DATA-CLEANING

- Language of your choice (e.g. Python, R, Ruby)
- Data-cleaning supported by your OS (e.g. Shell scripts)
- Data-cleaning tools (e.g. TextWrangler, **OpenRefine**,
GoogleFusion)

OPENREFINE (GOOGLEREFINE)

Google refine Private letters by W Colenso edit TK 2 Permalink Open... Export... Help

Facet / Filter Undo / Redo 28 615 records Extensions: Named-entity recognition ▾ Freebase ▾ RDF ▾

Show as: rows records Show: 5 10 25 50 records « first < previous 1 - 10 next > last »

All Date Date (year) Addressee Location Test 10 AlchemyAPI DBpedia Spotlight Zem...

★	1.	May 11, 1878	1878	D. Black	Napier	Mr D. Black Clyde, Wairoa, Dear Sir, On my return from the Country (from my last round as School Inspector) in the beginning of this month, I found your letter of 22nd April here, with many others, awaiting me. I could not however find time to reply to it last Tuesday's Mail, and I will now try to do so. With reference to Mr Luff's piece of land there at Wairoa, you say it would have been more satisfactory if you had been informed (by me) what difference there was between you and so, no doubt, it would, but I could not tell you any more: I do not know what price Mr Luff has set upon it. I know one thing, that he is a pretty good judge in all such matters, and that he is willing to sell it for what he considers a fair price. As you say, there is little chance (now that I am out of office) of my visiting Wairoa, otherwise I would go & look at the Land with you, & give you my opinion upon it, but, as that cannot be, I would recommend you to make a fair or good offer for it, and I will send it to Mr Luff: more than this I cannot say. I am &c (signed) W. Colenso P.S. I can well understand that while you are delaying about it, another may jump in & your chance is gone! W.C.	Mr Luff							Wairoa Choose ne...
★			1878	D. Black	Napier							Inspector Choose ne...		
★			1878	D. Black	Napier							Clyde F.C. Choose ne...		
★			1878	D. Black	Napier							Fair value Choose ne...		

OPENREFINE (GOOGLEREFINE)

GOOGLEFUSION

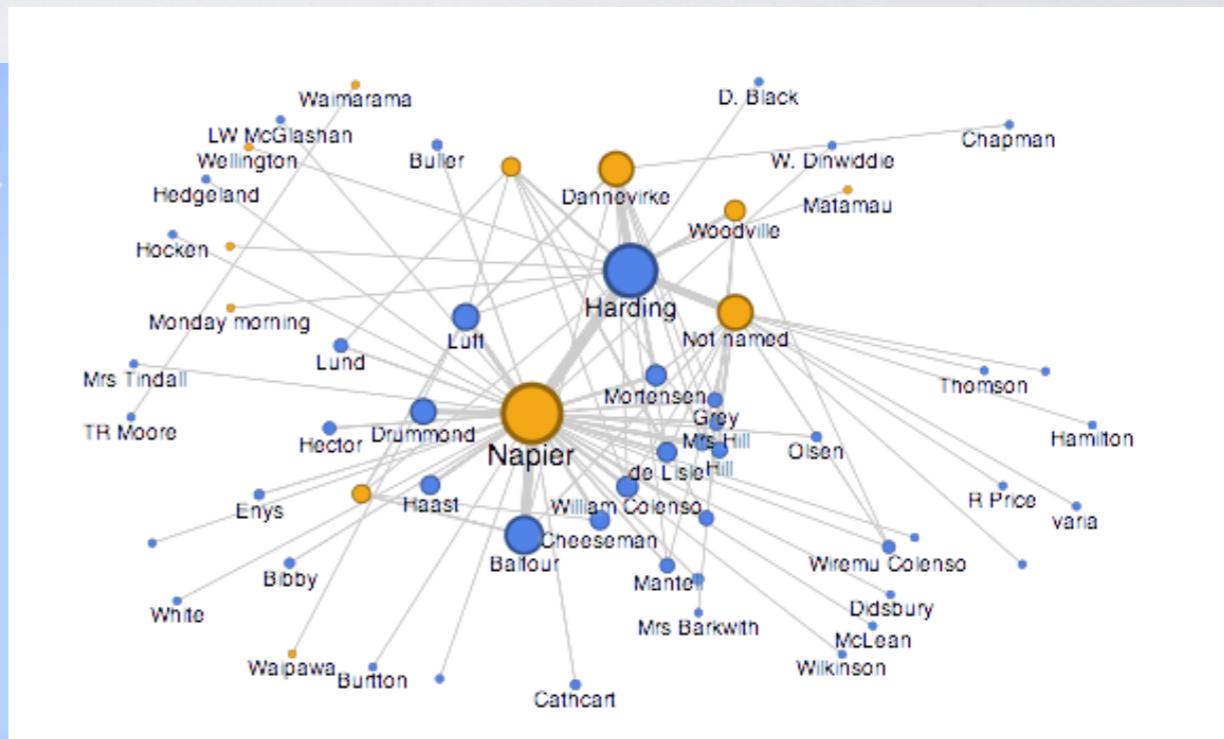
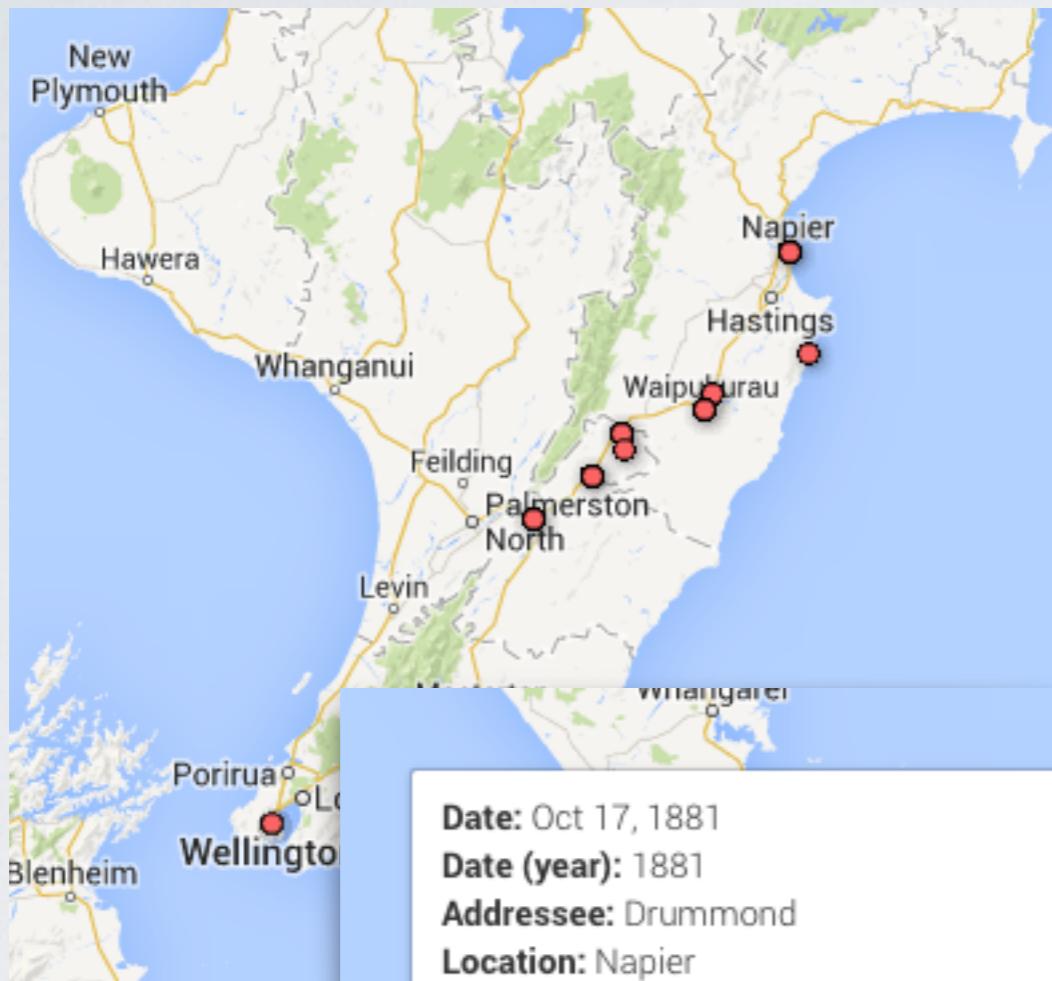
Private letters by W Colenso (edit TK)

Imported at Thu Nov 07 16:06:33 PST 2013 from Private-letters-by-W-Colenso-edit-TK.csv. Wai-te-a... [more >>](#)

Wai-te-ata Press - Edited on November 8, 2013

File Edit Tools Help				Data	Letters with metadata	Map of Colenso's private letters	Year - Addressee
Filter ▾ No filters applied							
1-100 of 615							
Date	Date (year)	Addressee	Location	File	Edit	Tools	Help
May 11, 1878	1878	D. Black		Data	Letters with metadata	Map of Colenso's private letters	Year - Addressee
May 11, 1878	1878	J.McCulloch		File	Edit	Tools	Help
May 21, 1878	1878	Luff		Data	Letters with metadata	Map of Colenso's private letters	Year - Addressee
Jun 12, 1878	1878	Luff		File	Edit	Tools	Help
Date: May 11, 1878 Date (year): 1878 Addressee: D. Black Location: Napier Letter Body: I Mr D. Black Clyde, Wairoa, Dear Sir, On my return from the Country (from my last round as School Inspector) in the beginning of this month, Äövñäel found your letter of 22nd April here, with many others, awaiting me. I could not however find time to reply to it last Tuesday, Äövñäts Mail, and I will now try to do so. With reference to Mr Luff, Äövñäts piece of land there at Wairoa, you say, Äövñä, Äövñä, Äövñäit would have been more satisfactory if you had been informed (by me) what difference there was between you., Äövñä, Äövñä, Äövñäand so, no doubt, it would, but I could not tell you any more: I do not know what price Mr Luff has set upon it. I know one thing, that he is a pretty good judge in all such matters, and that he is willing to sell it for what he considers a fair price. As you say, Äövñä, Äövñäthere is little chance (now that I am out of office) of my visiting Wairoa, otherwise I would go & look at the Land with you, & give you my opinion upon it., Äövñä, Äövñäbut, as that cannot be, I would recommend you to make a fair or good offer for it, and I will send it to Mr Luff: more than this I cannot say. I am &c (signed) W. Colenso P.S. I can well understand that while you are delaying about it, another may jump in & your chance is gone! W.C.				Date: May 11, 1878 Date (year): 1878 Addressee: J.McCulloch Location: Napier Letter Body: I Mr. John McCullough, Äövñä, Äövñä Sir, A short time ago, Mr Luff, Napier, has been received Rates on property for the year (seven shillings). As a friend of mine, Mr. Colenso, sent me this letter, and now send you enclosed a copy of the rates. Äövñä, Äövñäfor which please pay him. The rates have been better (more regular & reliable) than those of the said property is., Äövñä, ÄövñäIndeed I am wholly in the dark about the information: but, I suppose it is correct. Wm. Colenso.			
Date: May 21, 1878 Date (year): 1878 Addressee: Luff				Date: Jun 12, 1878 Date (year): 1878 Addressee: Luff			

GOOGLEFUSION



Date: Oct 17, 1881

Date (year): 1881

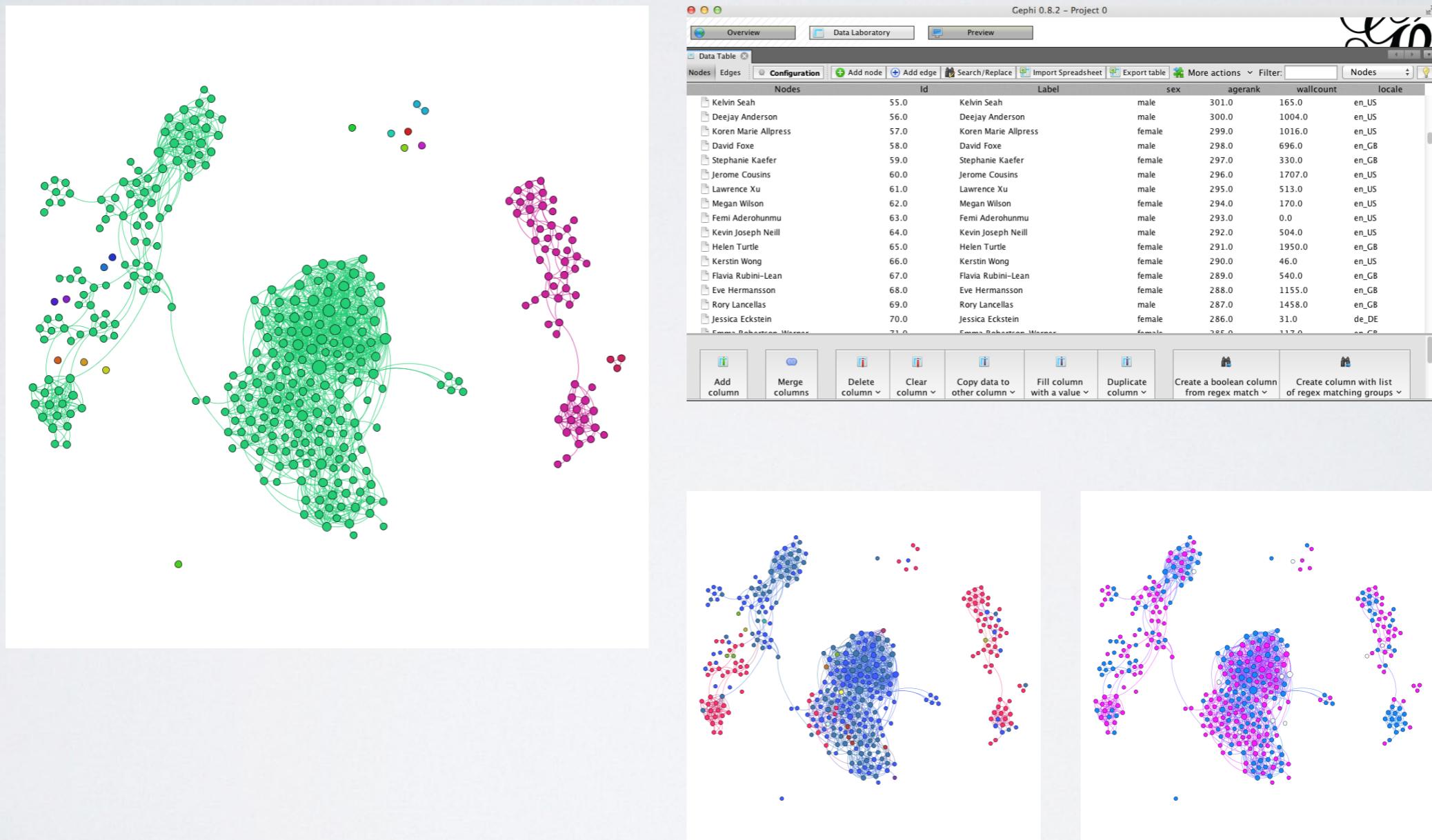
Addressee: Drummond

Location: Napier

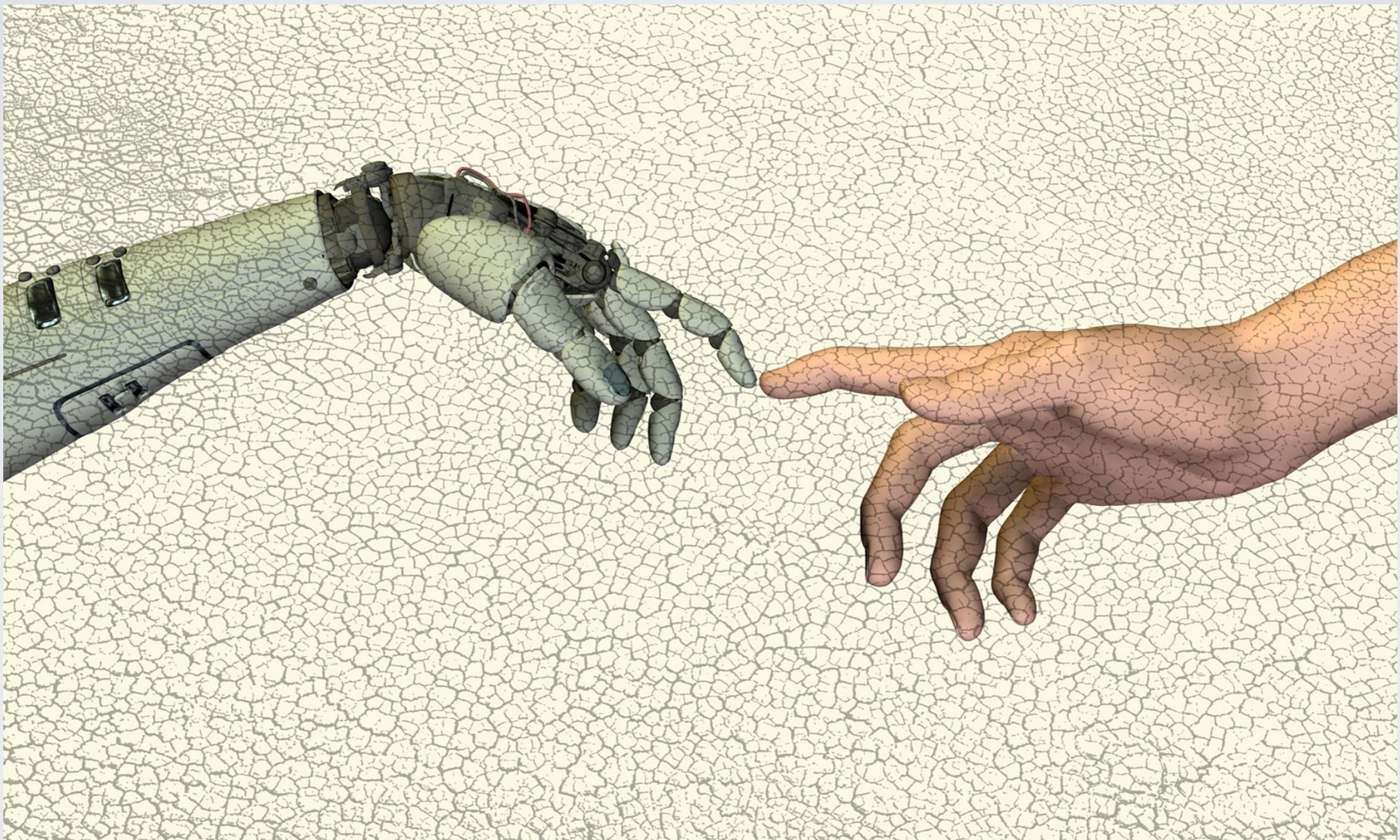
Letter Body: | Mr John Drummond || Dear Sir || No doubt you have been long expecting a line or two from me. I have been very busy (rather more so than I could have wished) having been considerably thrown back through Rheumatism, but this morning I cleared right off a lot of (Papers &c) to Dr Hector, and now, this afternoon & evening, I hope to finish reply-ing to some letters, which have been long waiting on my table. - - - - Äövñ\æ(I leave tomorrow for 40 mile Bush & shall return about end of month, all being well. I would have gone thither last week had it not been for the holy-days, as I never travel at such times). || Your letter of 19th Sept I duly received, & was pleased to find that you were then all well; Äövñ\æmay this find you so.. || I have just put up for you 6 Eng Papers, to be posted with this. They are rather old, & would have been sent to you before, only I had lent two of the lot, and one of them only came back on Sat last. Äövñ\æI had been personally 3 times after it. Those 2 Standards

GOOGLEFUSION

GEPHI



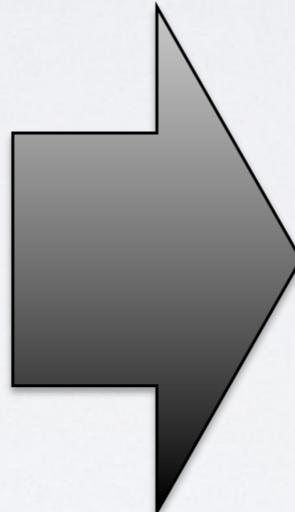
QUESTIONS?



PRACTICAL PART: OPEN-REFINE

DATA CLEANING

[{"id": "22702847", "updated_at": "2015-06-18T18:56:26.709+12:00", "created_at": "2012-04-21T07:23:26.000+12:00", "title": "Scott, Thomas, 1947- :[Christchurch earthquake] 25 February 2011", "description": "The cartoon shows a digger dredging through the rubble and digging up a red heart representing 'hope' (Tom Scott doesn't do colour so this is significant). A rescuer nearby yells 'Careful! It's still beating'. Context - on 22 February 2011 a 6.3 magnitude earthquake struck in Christchurch which has probably killed more than 200 people (at this point the number is still not known) and caused much more severe damage. There were many people trapped in collapsed buildings and it was apparent in only two or three days that in most cases they could not have survived but of course people still held out impossible hope.\nQuantity: 1 digital cartoon(s).\nPhysical Description: Image file - Jpeg", "content_partner": "Alexander Turnbull Library", "category": "Images", "creator": "Not specified", "dc_type": "Item", "dnz_type": "Artwork", "date": "2011-01-01 12:00:00 UTC", "source_url": "http://api.digitalnz.org/records/22702847/source", "collection": "Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "TAPUHI": "Drawings and Prints Collection", "New Zealand Cartoon Archive", "CEISMIC", "alternate_title": "", "additional_description": "", "display_content_partner": "Alexander Turnbull Library", "collection_title": "Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "TAPUHI": "Drawings and Prints Collection", "New Zealand Cartoon Archive", "CEISMIC", "display_collection": "TAPUHI", "primary_collection": "TAPUHI", "contributing_partner": "", "contributor": "", "copyright": "All rights reserved", "citation": "", "credit_creator": null, "language": "en", "provenance": null, "publisher": null, "rights": "Please check copyright", "usage": "All rights reserved", "source": null, "tag": "CEISMIC", "thesis_level": null, "holding": null, "library_collection": "Drawings and Prints Collection", "New Zealand Cartoon Archive", "shelf_location": "DCDL-0017168", "eprints_type": null, "text": null, "fulltext": null, "format": "Digital images", "Cartoons (Commentary)": 1, "digital_cartoon": "Single art work", "Image_file_Jpeg": null, "dc_identifier": "ndha:E3303922", "tap1410005", "urn:nbn:nz:wtu:DCDL-0017168", "DCDL-0017168", "display_date": "2011", "published_date": null, "syndication_date": "2014-08-04T15:29:35.905+12:00", "landing_url": "http://natlib.govt.nz/records/22702847", "large_thumbnail_url": "http://ndhadeliver.natlib.govt.nz/NLNZStreamGate/get?dps_pid=IE3303922", "rights_url": null, "thumbnail_url": "http://ndhadeliver.natlib.govt.nz/delivery/DeliveryManagerServlet?dps_pid=IE3303922\u0026dps_func=thumbnail", "origin_url": "http://tapuhi.natlib.govt.nz/cgi-bin/spydus/NAV/GLOBAL/OPHDR/1/1410005", "metadata_url": null, "object_url": null, "has_version": null, "license": null, "relation": "Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "DC_Group-0025": "tap1024981", "is_part_of": "Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]", "DC_Group-0025": "DC-Group-0025", "tap1024981": "is_replaced_by": null, "replaces": null, "is_required_by": null, "is_version_of": null, "table_of_contents": null, "is_commercial_use": null, "atl_free_download": null, "atl_physical_viewability": null, "atl_purchasable": true, "atl_purchasable_download": true, "atl_location_code": null, "atl_usage_code": null, "anzsrc_code": null, "marsden_code": null, "subject": "Hope", "Christchurch Earthquake, N.Z., 2011", "Earthquakes", "New Zealand", "Canterbury Region", "Search and rescue operations", "Christchurch City", "coverage": null, "attachments": [{"id": "5556b31c646e7a5bbd396500", "aspect_ratio": null, "date": null, "dc_identifier": "IE3303922", "dc_type": null, "description": null, "display_date": null, "file_size": null, "file_type": null, "large_thumbnail_url": "http://ndhadeliver.natlib.govt.nz/NLNZStreamGate/get?dps_pid=IE3303922", "name": "25feb11.jpg", "ndha_rights": 100, "thumbnail_url": "http://ndhadeliver.natlib.govt.nz/delivery/DeliveryManagerServlet?dps_pid=IE3303922\u0026dps_func=thumbnail", "title": null, "url": null}], "authorities": [{"id": "55826b9a646e7a095201d628", "authority_id": 1137550, "name": "subject_authority", "role": null, "text": "Hope"}, {"id": "55826b9a646e7a095201d629", "authority_id": 1411976, "name": "subject_authority", "role": null, "text": "Christchurch Earthquake, N.Z., 2011"}, {"id": "55826b9a646e7a095201d62a", "authority_id": 144081, "name": "subject_authority", "role": null, "text": "Earthquakes - New Zealand - Canterbury"}]



Cartoonist

Title

Date

URL

Publisher

Description

Name authorities

Subject authorities

RESTRUCTURING

- Cartoonist
- Title
- Date
- URL
- Publisher
- Description
- Authorities

I. INSTALL OPENREFINE

<http://openrefine.org/download.html>

II. DOWNLOAD EXAMPLE DATASET FROM GITHUB

git pull
or

git clone <https://github.com/ThomasK81/IntroDH.git>

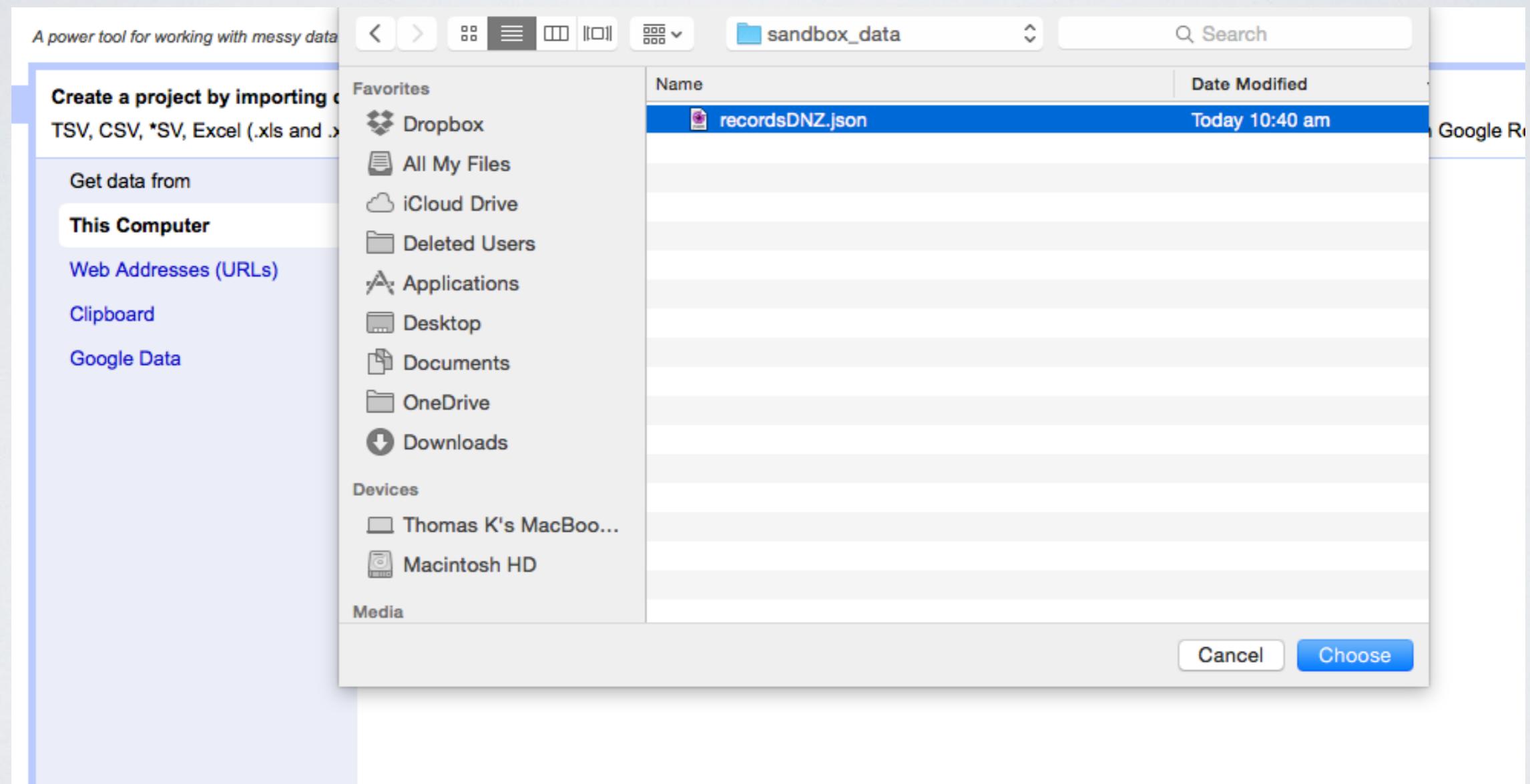
or

[http://api.digitalnz.org/v3/records.json?api_key=LUq9-soDzWWWhShuy3XhU&and\[category\]](http://api.digitalnz.org/v3/records.json?api_key=LUq9-soDzWWWhShuy3XhU&and[category])

[\[\]&Images&and\[collection\]](#) [\[\]&New+Zealand+Cartoon](#)

[+Archive&fields=id,shelf_location,source_url,title,description,](#)
[collection&per_page=100&page=3](#)

III. OPENREFINE AND IMPORT DATA SET



III. OPENREFINE AND IMPORT DATA SET

« Start Over Configure Parsing Options Project name recordsDNZ json

Click on the first JSON {} node corresponding to the first record to load.

```
{  
  search: {  
    results: [  
      {  
        published_date: [],  
        holding: [],  
        tag: [  
          CEISMIC  
        ],  
        rights: Please check copyright,  
        primary_collection: [  
          TAPUHI  
        ],  
        display_collection: TAPUHI,  
        ...  
      }  
    ]  
  }  
}
```

Parse data as Pick Re...

JSON files Load at most 0 record(s) of data

Line-based text files Store file source (file names, URLs) in each row

CSV / TSV / separator-based files

Fixed-width field text files

IV. ADD AUTHORITIES

Create column Authorities at index 1 by fetching URLs
based on column _ - id using expression grel:"<http://api.digitalnz.org/v3/records/>" + value + ".json?
api_key=LUq9-
soDzWWhShuy3XhU&fields=authorities"

PRUNING

- Double-ups from Canterbury & Christchurch
- Remove Photographs
- Remove DPP and Ephemera

CREATE CSV FILE WITH THIS INFORMATION

- Cartoonist
- Title
- Date
- URL
- Publisher
- Description
- Authorities

EXTRACT CARTOONIST'S NAME

- Cartoonist, title and date all in one field
- But, cartoonist is not always in title
 - As cartoonist
 - As creator
 - Collection_root

EXTRACT THE DATE

"id":22702847,"updated_at":"**2015-06-18T18:56:26.709+12:00**","created_at":"**2012-04-21T07:23:26.000+12:00**","title":"Scott, Thomas, 1947- :[Christchurch earthquake] **25 February 2011**","description":"The cartoon shows a digger dredging through the rubble and digging up a red heart representing 'hope' (Tom Scott doesn't do colour so this is significant). A rescuer nearby yells 'Careful! It's still beating'. Context - on 22 February 2011 a 6.3 magnitude earthquake struck in Christchurch which has probably killed more than 200 people (at this point the number is still not known) and caused much more severe damage. There were many people trapped in collapsed buildings and it was apparent in only two or three days that in most cases they could not have survived but of course people still held out impossible hope.\nQuantity: 1 digital cartoon(s).\nPhysical Description: Image file - Jpeg","content_partner":["Alexander Turnbull Library"],"category":["Images"],"creator":["Not specified"],"dc_type":["Item"],"dnz_type":"**Artwork**","date":["**2011-01-01 12:00:00 UTC**"],"source_url":"http://api.digitalnz.org/records/22702847/source","collection":["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]","TAPUHI","Drawings and Prints Collection","New Zealand Cartoon Archive","CEISMIC"],"alternate_title":[],"additional_description":[],"display_content_partner":"Alexander Turnbull Library","collection_title":["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]","TAPUHI","Drawings and Prints Collection","New Zealand Cartoon Archive","CEISMIC"],"display_collection":"TAPUHI","primary_collection":["TAPUHI"],"contributing_partner":[],"contributor":[],"copyright":["All rights reserved"],"citation":[],"credit_creator":null,"language":["en"],"provenance":null,"publisher":[],"rights":"Please check copyright","usage":["All rights reserved"],"source":[],"tag":["CEISMIC"],"thesis_level":null,"holding":[],"library_collection":["Drawings and Prints Collection","New Zealand Cartoon Archive"],"shelf_location":"DCDL-0017168","eprints_type":[],"text":null,"fulltext":null,"format":["Digital images","Cartoons (Commentary)"],"1 digital cartoon(s)","Single art work","Image file - Jpeg"],"dc_identifier":["ndha:IE3303922","tap:1410005","urn:nbn:nz:wtu:DCDL-0017168","DCDL-0017168"],"display_date":"**2011**","published_date":[],"syndication_date":"**2014-08-04T15:29:35.905+12:00**","

EXTRACT THE DATE

"id":22702847,"updated_at":"**2015-06-18T18:56:26.709+12:00**","created_at":"**2012-04-21T07:23:26.000+12:00**","title":"Scott, Thomas, 1947- :[Christchurch earthquake] **25 February 2011**","description":"The cartoon shows a digger dredging through the rubble and digging up a red heart representing 'hope' (Tom Scott doesn't do colour so this is significant). A rescuer nearby yells 'Careful! It's still beating'. Context - on 22 February 2011 a 6.3 magnitude earthquake struck in Christchurch which has probably killed more than 200 people (at this point the number is still not known) and caused much more severe damage. There were many people trapped in collapsed buildings and it was apparent in only two or three days that in most cases they could not have survived but of course people still held out impossible hope.\nQuantity: 1 digital cartoon(s).\nPhysical Description: Image file - Jpeg","content_partner":["Alexander Turnbull Library"],"category":["Images"],"creator":["Not specified"],"dc_type":["Item"],"dnz_type":"**Artwork**","date":["**2011-01-01 12:00:00 UTC**"],"source_url":"http://api.digitalnz.org/records/22702847/source","collection":["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]","TAPUHI","Drawings and Prints Collection","New Zealand Cartoon Archive","CEISMIC"],"alternate_title":[],"additional_description":[],"display_content_partner":"Alexander Turnbull Library","collection_title":["Scott, Thomas, 1947- :[Digital cartoons published from 2003 onward in the Dominion Post]","TAPUHI","Drawings and Prints Collection","New Zealand Cartoon Archive","CEISMIC"],"display_collection":"TAPUHI","primary_collection":["TAPUHI"],"contributing_partner":[],"contributor":[],"copyright":["All rights reserved"],"citation":[],"credit_creator":null,"language":["en"],"provenance":null,"publisher":[],"rights":"Please check copyright","usage":["All rights reserved"],"source":[],"tag":["CEISMIC"],"thesis_level":null,"holding":[],"library_collection":["Drawings and Prints Collection","New Zealand Cartoon Archive"],"shelf_location":"DCDL-0017168","eprints_type":[],"text":null,"fulltext":null,"format":["Digital images","Cartoons (Commentary)"],"1 digital cartoon(s)","Single art work","Image file - Jpeg"],"dc_identifier":["ndha:IE3303922","tap:1410005","urn:nbn:nz:wtu:DCDL-0017168","DCDL-0017168"],"display_date":"**2011**","published_date":[],"syndication_date":"**2014-08-04T15:29:35.905+12:00**","

"Create column Extracted Date at index 1 based on
column title using expression gREL:
value.partition([0-9]{1,2}+\\"s+\\"S+\\"s+[0-9]{4})/[1]

CREATE CSV FILE WITH THIS INFORMATION

- Cartoonist
- Title
- Date
- URL
- Publisher
- Description
- Authorities

TASK
EXPORT CSV TO
GOOGLEFUSION
AND
GITHUB