

# Mailing Lists - Gmane

- Crawl the archive of a mailing list
- Do some analysis / cleanup
- Visualize the data as word cloud and lines



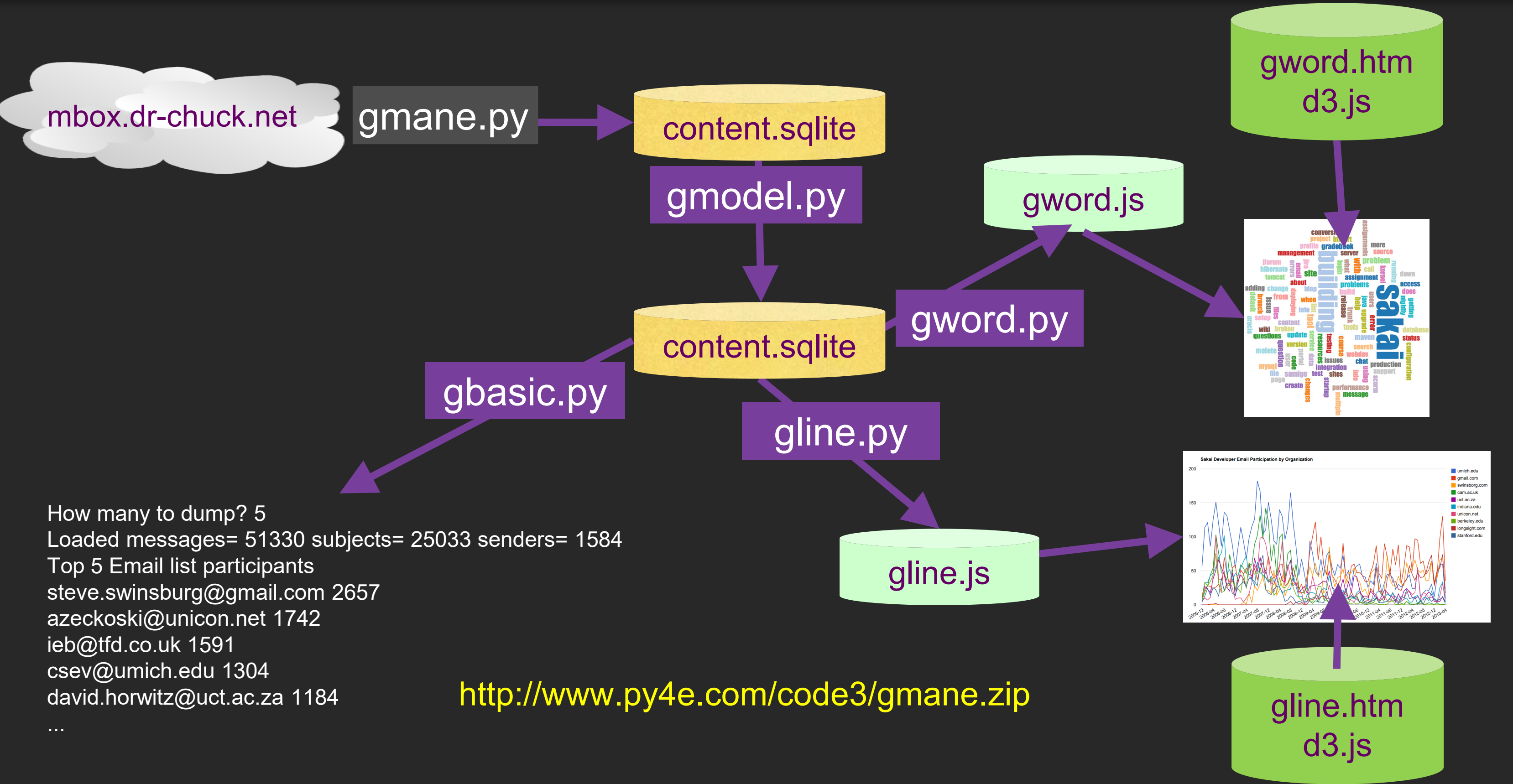
<http://www.py4e.com/code3/gmane.zip>

# Warning: This Dataset is > 1GB

- Sadly the original source of this data ([gmane.org](http://gmane.org)) has shut down
- We made a copy of a subset of the data before it was shut down

Use this for your testing:

<https://mbox.dr-chuck.net/sakai.devel/4/5>



# Summary

- Combining all of the techniques in the class, we can make data management system that can pull data (restarting as necessary), clean it up, and visualize it



## Acknowledgements / Contributions



These slides are Copyright 2010- Charles R. Severance ([www.dr-chuck.com](http://www.dr-chuck.com)) of the University of Michigan School of Information and [open.umich.edu](http://open.umich.edu) and made available under a Creative Commons Attribution 4.0 License. Please maintain this last slide in all copies of the document to comply with the attribution requirements of the license. If you make a change, feel free to add your name and organization to the list of contributors on this page as you republish the materials.

Initial Development: Charles Severance, University of Michigan School of Information

... Insert new Contributors here