

2016-09-14 Meeting notes

Date

14 Sep 2016

Attendees

- [Unknown User \(a-craig\)](#)
- [Sandeep Puthanveetil Satheesan](#)
- [Paul Rodriguez](#)
- [Elizabeth Wuerffel](#)
- [Jeff Will](#)

Goals

Discussion items

Time	Item	Who	Notes
	OCR	Liz	OCR not working well, even with very text-based images.
		Paul	171k images. 86k have tiff file "largest." 84k have tiff file called "larger." Difference in resolution. Jeff: need to develop ranges to process
		Jeff	Would be nice to know how long it would take to run full corpus to make case for using supercomputers

Action items

- ☒ [Sandeep Puthanveetil Satheesan](#): Can we improve the OCR through thresholding or anything else to enhance?
- ☐ [Paul Rodriguez](#): Investigating different resolutions sizes based on file name "large" and "larger," etc.