REQUEST FOR RECORDS DISPOSITION AUTHORITY	JOB NUMBER N1-064-05-6
To: NATIONAL ARCHIVES & RECORDS ADMINISTRATION 8601 ADELPHI ROAD COLLEGE PARK, MD 20740-6001	Date received 2/14/2005
1 FROM (Agency or establishment) National Archives & Records Administration	NOTIFICATION TO AGENCY
2 MAJOR SUBDIVISION Modern Records Program	In accordance with the provisions of 44 U S C 3303a, the disposition request, including amendments, is approved except for items that may be marked "disposition not approved" or "withdrawn" in column 10
3 MINOR SUBDIVISION	

4 NAME OF PERSON WITH WHOM TO CONFER	5 TELEPHONE NUMBER	DATE	ARCHIVIST OF THE UNITED STATES
John Kane	301-837-1604	3/21/05	Allen Weinstein

6 AGENCY CERTIFICATION

I hereby certify that I am authorized to act for this agency in matters pertaining to the disposition of its records and that the records proposed for disposal on the attached 1 page(s) are not needed now for the business for this agency or will not be needed after the retention periods specified, and that written concurrence from the General Accounting Office, under the provisions of Title 8 of the GAO Manual for Guidance of Federal Agencies,

☑ is not required ☐ is attached, or ☐ has been requested

DATE	SIGNATURE OF AGENCY REPRESENTATIVE	TITLE
FEB 14 2005	Richard W Marcus	NARA Records Officer

7 ITEM NO	8 DESCRIPTION OF ITEM AND PROPOSED DISPOSITION	9 GRS OR SUPERSEDED JOB CITATION	10 ACTION TAKEN (NARA USE ONLY)
	Federal Agencies Web Harvest Content Files See attached page(s)		

3/22/05 Copies to Agmt / NWME

Federal Agencies Web Harvest Content Files

At the conclusion of each Administration, NARA conducts a project to capture and preserve a government-wide snapshot of federal public websites (NARA conducted such an effort at the conclusion of the Clinton Administration and first Bush term.)

Master and back-up.

Disposition: PERMANENT. Transfer to the National Archives upon approval of this schedule

Background information
The End-of-Presidential-Term web harvest of 2004 represents web content harvested between 10/14/04 and 11/19/04 using the Heritrix web harvester and a seed list of 1658 unrestricted Federal web site URLs provided by GSA (DotGov Registry) and DISA at the 2nd or deeper domain levels The web content was captured in the ARC File Format and provided to NARA in that format. It was also provided to NARA as content converted back to its native file formats (as they appeared on the web. .i e. HTML, PDF etc)

ARC is the open source web preservation format created by Internet Archives and used as a standard format by the International Internet Preservation Consortium Each ARC file consists of a sequence of compressed web documents with a header line of metadata. The header is followed by the actual data returned by the web protocol. The main fields of the header line are the URI of the document, the timestamp of when the data were acquired and the size of the data block following the header