

Recommendations of the HDMC Analytics Working Group 2010

The Heliophysics Data and Model Consortium (HDMC) established the Analytics working group in June 2010.

Goals

The goal of the Analytics group is to determine which analytics of system activities are necessary to provide insights into the effectiveness and usefulness of the elements in the HDMC. The elements of the HDMC include Virtual Observatories, resident archives, active archives, value-added services, metadata inventories, and data movement. A secondary goal is to recommend an approach for collecting the metrics necessary for generating the analytics and a method for generating the desired analytics.

Definitions

PPI: Personally Identifiable Information (PPI) is Information that can be used to uniquely identify, contact, or locate a single person or can be used with other sources to uniquely identify a single individual. Examples are: name, SSN, biometric records, an identifying number or code, or other identifying particular assigned to the individual

- PII excludes strictly business contact information (such as work e-mail address, office location, and office telephone number) [NPR 1382.1, Sec 1.1.2]

Element: A operating unit or component of the system which has a distinct function.

For the HDMC this can be a Virtual Observatory, Resident Archive, service, or value-added capability.

Reporting Requirements

Elements of the HDMC should report metrics which can provide insight into effectiveness of achieving the goals of the element. For the HDMC the general metrics to collection are:

- Number of visits or uses of an element.
- Bytes transferred.
- Country of origin of requests.
- References to the element in publications.
- References to managed items (data) in publications.

To support mission or data provider reporting it must be possible to provide access and transfer metrics summarized by the unique resource identifier.

Governing Requirements

As a NASA funded activity the HDMC and the participating elements of the HDMC (VxO, Resident Archives, etc.) are governed by NASA policies and directives. The policies that relate to the collection of metrics and generation of analytics for web sites and web services include:

NASA Privacy Policy , NPD 1382.17H, effective: June 24, 2009 - June 24, 2014

http://nodis3.gsfc.nasa.gov/npg_img/N_PD_1382_017H_/N_PD_1382_017H_main.pdf

NASA Privacy Procedural Requirements, NPR 1382.1, effective: August 10, 2007-August 10, 2012

http://nodis3.gsfc.nasa.gov/npg_img/N_PR_1382_0001_/N_PR_1382_0001_.pdf

NASA Privacy Procedural Requirements – Chapter 4: NASA Web Site Privacy, NPR 1382.1, effective: August 10, 2007-August 10, 2012

http://nodis3.gsfc.nasa.gov/npg_img/N_PR_1382_0001_/N_PR_1382_0001_Chapter4.pdf

NASA Privacy Procedural Requirements – Appendix G: Web Privacy Guidance, NPR 1382.1, effective: August 10, 2007-August 10, 2012

http://nodis3.gsfc.nasa.gov/npg_img/N_PR_1382_0001_/N_PR_1382_0001_AppendixG.pdf

The Federal laws that impact NASA's collection and management of personally identifiable information (PII) include the Privacy Act of 1974 (hereinafter referred to as the Privacy Act), the Children's Online Privacy Protection Act of 1998 (COPPA), the E-Government Act of 2002 (E Gov Act), the Federal Information Security Management Act of 2002 (FISMA), and the Paperwork Reduction Act of 1995 (PRA). The intent of these laws is embodied in the referenced NASA Policy and Procedural Requirements documents.

Findings

The HDMC Analytics Working Group has determined the following:

1. The collection of metrics must comply with the NASA privacy policies.
2. A web site must have a "Web Privacy Policy" that specifies what information NASA collects automatically (e.g., user's internet protocol (IP) address, location, and time of visit) and identifies the use for which it is collected (i.e., site management or security purposes). [NPR 1382.1, 4.3.1.e]
3. The "Web Privacy Policy" be Clearly labeled "Privacy Policy and Important Notices." on the web site. [NPR 1382.1, 4.3.3.a]
4. Web sites should not use persistent COOKIES on public NASA Web sites unless there is a compelling reason to do so. The conditions under which persistent cookies may be used are detailed in [NPR 1382.1, 4.4.5]
5. Information collected using cookies (see Chapter 4) and Web server logs is not covered by the Privacy Act unless the information collected is routinely retrieved by a personal identifier. [NPR 1382.1, G.1.c]

Recommendations

1. Each identified element in the HDMC will report the following for each designated reporting period:
 - Number of visits or uses of an element.
 - Raw
 - Excluding known bots
 - By country of origin
 - Bytes transferred.
 - Raw
 - Excluding known bots
 - By country of origin.
 - Managed resources
 - Access (or visit)
 - Bytes transferred
 - References to the element in publications.
 - References to managed resources (data) in publications.

- By Unique identifier

Visits and byte counts may be determined using third party tools or services provided that the information transferred does not contain any personal identifiers. The IP address of a system is not considered PII.

2. Elements are to provide open access whenever contractual possible to avoid constraints which are imposed when personal identification information (PII) is collected.
3. Establish reporting timeframes. Weekly reporting will allow a flexible resolution for the generation of monthly, quarterly, and annual reports.
4. Establish a common method of reporting. Options include e-mailed spreadsheets with a common layout, a web form, or a web logging service (Google Analytics is an example).