NOAA Best Practice for the Use of External Data

Version 1.0
2013 February 26

NOAA Environmental Data Management Committee

Table of Contents

Revision History

Version	Date	Name	Description
0.1	2012 Aug 15	James W. Sargent	First Draft circulated for EDMC review.
0.2	2012 Sep 12	Jeff de La Beaujardière, James W. Sargent,	Second draft for EDMC review.
0.3	2012 Sep 27	Jeff de La Beaujardière	Third draft for NOSC and CIO Council review. Definitions moved into body of document. Related Policy section improved.
0.4	2012 Nov 15	"	Fourth draft for NEP/NEC review. Editorial changes only.
1.0	2013 Feb 26	"	Version 1.0 for presentation to SAB. No changes from v0.4

Acknowledgements

1. Purpose

This document defines a Best Practice for US National Oceanic and Atmospheric Administration (NOAA) Programs and Projects regarding the use of data from external (non-NOAA) sources. The best practice consists of answering the questions in the attached External Data Usage Worksheet (Appendix A) to the satisfaction of the Program or Project's responsible NOAA official.

The External Data Usage Worksheet (Appendix A) includes questions in the following categories:
- Description
- Purpose
- Quality
- Reliability
- Terms of Use
- IT Systems and Security
- Life-Cycle Cost
- Metadata
- Accessibility
- Archiving
- Formal Agreements

The worksheet should be used as a checklist to assess whether to accept external data for the intended purpose and to ensure consideration is given to the need for discovery, access, and long-term preservation within NOAA. Each question should be assessed for relevance to the project and its broader agency context, the specific data in question, and the intended use of the data. Issues raised by the questions should be understood and the questions answered to the satisfaction of the project and any appropriate authorities. The Worksheet should be considered independently for each data source and, if multiple and disparate datasets are involved, for each dataset from each source.

Appendix B lists related NOAA policies and procedures.

2. Rationale and Background

NOAA's core activities include collecting, processing, and integrating environmental data to understand, forecast, and respond to Earth's terrestrial, atmospheric, and oceanic phenomena. NOAA information products, analyses, and decisions may require data from both NOAA and external sources. NOAA's credibility and reliability are at stake if NOAA uses external data that are inaccurate, unreliable, or not within NOAA's legal right to use. This Best Practice is therefore intended to ensure that external data are used judiciously.

The NOAA Science Advisory Board (SAB) Data Archive and Access Requirements Working Group (DAARWG) presented a report entitled "Assessing the Use of Data from Non-NOAA Sources" to the SAB in November, 2011. The report recommended developing and implementing NOAA-wide guidelines and identified specific elements to be included therein. The SAB endorsed the recommendation and passed it on to NOAA leadership for a response. Subsequently, the Environmental Data Management Committee (EDMC) was tasked to develop the response to be delivered to the SAB at their Spring 2013 meeting. This document constitutes the NOAA response to the SAB recommendation.

44 3. Scope

45 This NOAA Best Practice applies to all NOAA projects or programs that wish to obtain environmental
46 data or socio-economic data from external sources, especially if such data would be used in products
47 that receive the NOAA imprimatur or in contexts that may affect life, property, or highly influential
48 scientific assessments.

49 **Environmental Data** are recorded and derived observations and measurements of the physical,
50 chemical, biological, geological, and geophysical properties and conditions of the oceans, atmosphere,
51 space environment, sun, and solid earth, as well as correlative data, such as socio-economic data,
52 related documentation, and metadata. (From NOAA Administrative Order (NAO) 212-15, *Management*
53 *of Environmental and Geospatial Data and Information* (2003) [1].

54 **Socio-economic data** are observations and measurements of the ways the environment affects humans
55 economically, socially, or culturally.

56 **External sources** include, but are not limited to:

57 • Other federal agencies
58 • State, local, or tribal governments
59 • NOAA grantees or contractors
60 • Non-governmental organizations (NGOs)
61 • Commercial organizations (whether for-profit or not)
62 • Agencies of other national governments
63 • Research and educational institutions
64 • General public (e.g., "crowd-sourcing")

65 **NOAA products** include, but are not limited to:

66 • data records and products derived from satellite and other observations
67 • weather forecasts
68 • assessments or forecasts of hazardous conditions, ecosystems, or climate
69 • thematic portals that provide access to both NOAA and external data sources.

70
71 This Best Practice should be followed prior to new uses of external data. Existing projects that already
72 use external data should review the Worksheet in Appendix A to ensure they have considered the
73 matters therein. In emergency-response situations it may be necessary to streamline the process
74 initially, but personnel should at least familiarize themselves with these guidelines in advance to avoid
75 inadvertent misuse of external data, and should perform a fuller assessment as soon as the situation
76 permits.

77
78 Enforcement and monitoring of this NOAA Best Practice is at the discretion of NOAA Line and Staff
79 Offices and should be appropriate to the level of sensitivity or risk of the programs and projects
80 concerned. The National Weather Service (NWS) has previously established NWS Policy Directive 1-12
81 and NWS Instruction 1-1201, which focus on rights and restrictions regarding data use and
82 redistribution.

83

---

[1] See Appendix C for references.

Appendix A. External Data Usage Worksheet

84

85 1. <u>Description</u>: Identify the external data, information, or products to be used, and their sources.

86 2. <u>Purpose</u>:

87   2.1. How NOAA will use the data?

88   2.2. Are related data available at NOAA to meet the requirements? Consult the NOAA Observing
89   Systems Architecture (NOSA) database maintained by the Technology Planning and Integration
90   for Observations (TPIO) program at https://www.nosc.noaa.gov/tpio/.

91   2.3. Is there an objective requirement for these external data? Has that requirement been
92   documented in the NOAA Consolidated Observing Requirements List (CORL) maintained by
93   TPIO (https://www.nosc.noaa.gov/tpio/)? Need should be the paramount factor in determining
94   whether to use external data.

95   2.4. Is the data source a critical or supplementary source? Are there alternative external sources?

96   2.5. Are the external data to be used temporarily, sporadically or continually?

97   2.6. Do emergency conditions justify the use of external data without a complete assessment based
98   on this Worksheet?

99 3. <u>Quality</u>:

100   3.1. Are the external data of known quality and of good quality?

101   3.2. Is the data quality documented in the metadata? (See also Metadata questions, below.)

102   3.3. What quality control procedures are being followed?

103   3.4. Do the data comply with the Information Quality Act (IQA), which applies to Federal agencies?

104   3.5. Will use of these data impact NOAA's ability to comply with the IQA? (See Appendix B for
105   information about NOAA IQA Guidelines.)

106   3.6. How will changes in the quality of an external data stream be detected?

107   3.7. How robust are the science algorithms?

108 4. <u>Reliability</u>:

109   4.1. Is the external data source reliable? With an external data stream, risks may be associated with
110   problems in the source or network. These can lead to loss of accuracy and reliability in NOAA
111   products.

112   4.2. For operational use, will the data source be operationally reliable (highly available and
113   redundant)?

114   4.3. Does NOAA use require that the data come from a certified source? If so, what constitutes
115   appropriate certification for these data and their intended use?

116   4.4. How likely is a sudden loss of the external data stream through network or data-source
117   problems?

118   4.5. What would be the impact to NOAA's mission if there were a sudden loss of this external data
119   stream?

120 5. <u>Terms of Use</u>:

121   5.1. Will the external data be made available to the public by NOAA? It is NOAA policy to make data
122   in its possession available to the public without restriction, to the extent practical within legal
123   and resource constraints. NOAA's commitment to open and unrestricted data may be affected
124   whenever it agrees to terms of use for data obtained from external parties that include
125   restrictions on NOAA's distribution of the data. NOAA use of external data does not constitute
126   endorsement of the provider's products. NOAA may describe its use of external data in its
127   derivative products but is not required to do so.

128   5.2. What are the restrictions or conditions, if any, on NOAA's use of the data? Will NOAA be
129   exempt from these restrictions when:

130		5.2.1.NOAA products derived from the data it receives do not contain sufficient information to
131		allow the original content of the data provided to be readily retrieved;
132		5.2.2.NOAA deems it appropriate to distribute the data in order to protect life and property in
133		situations that would constitute and emergency;
134		5.2.3.The data are received from a third party or have become a matter of public knowledge;
135		5.2.4.The data are redistributed to other federal agencies when access to the data is needed to
136		fulfill their respective missions;
137		5.2.5.The data are cited by NOAA incidentally in a NOAA product or publication;
138		5.2.6.Exemption is required by applicable laws.
139		5.3. Do the data include personally identifiable information (PII)? If so, does NOAA have the means
140		to safeguard the PII?
141		5.4. Do the data have a national security classification? National Security Information is information
142		that has been determined, pursuant to Executive Order 12958 or any predecessor order, to
143		require protection against unauthorized disclosure.
144		5.5. Does the provider of the external data want acknowledgement? If so, how will NOAA
145		acknowledge the data provider?
146	6.	Information technology (IT) systems and security:
147		6.1. What NOAA IT systems will be involved in the data exchange?
148		6.2. Do NOAA systems have capacity to receive the data and comply with the terms of use?
149		6.3. Will additional hardware or software be needed?
150		6.4. Have the NOAA systems owners been consulted?
151		6.5. What procedures are in place for security quarantine and anti-virus/anti-malware scanning of
152		data and external media before use of external data by NOAA?
153		6.6. What controls are used to protect data in transit from the source to ensure data integrity and
154		to prevent interception, loss, and modification?
155		6.7. What controls are used to authenticate the data provider to ensure that the data are obtained
156		from the authoritative source of the data?
157		6.8. Does the connection to the external provider require an Interconnection Security Agreement?
158	7.	Life-Cycle Cost:
159		7.1. What is the life-cycle cost of using the external data? Even "free" data may engender long-
160		term life-cycle costs and obligations. NOAA projects should estimate the cost relative to the
161		perceived benefit before obtaining data from external sources. Examples of cost include:
162		7.1.1.Purchase price;
163		7.1.2.Effort required to ingest, transform, or adapt the information for NOAA's purposes;
164		7.1.3.Cost of archival storage or long-term stewardship (see also Archiving questions, below);
165		7.1.4.Cost of ongoing data reprocessing, recalibrations, and version control;
166		7.1.5.Cost associated with complying with restrictions on use of the data.
167		7.1.6.Increased purchased price to allow unrestricted use of the data.
168		7.2. How do the total costs compare with the cost of establishing a new observing capability at
169		NOAA?
170	8.	Metadata:
171		8.1. Are the data adequately documented? Data documentation (metadata) must be adequate for
172		immediate and future use. The sources of data and the procedures used to develop products
173		and services from the data should, as much as practical, be transparently evident. Metadata
174		should be close to the data and bound to the data, if possible. The limitations of saved copies
175		of datasets to avoid misuses should be documented.
176		8.2. Is the provenance known and documented?
177		8.3. Have the data been assigned a persistent identifier? Will the data be properly cited when used?

178 9. <u>Accessibility:</u>
179   9.1. Will the data be appropriately accessible? For ongoing retrieval of data from an external
180   source, will data be readily accessible via a standardized protocol in a well-known format?
181   9.2. Does NOAA need to provide access to the data to other users?
182 10. <u>Archiving</u>:
183   10.1.   Are the data preserved in a long-term archive by the external source?
184   10.2.   If not, what are NOAA's obligation for long-term archiving of the data?
185   10.3.   If appropriate, has a Request-to-Archive process been initiated with a NOAA Data Center
186   with a follow-on submission agreement to support this external data source? (See *NOAA*
187   *Procedure for Scientific Records Appraisal and Archive Approval*, referenced in Appendix B).
188 11. <u>Formal Agreements</u>:
189   11.1.   Will the use of external data be subject to a formal agreement between NOAA and the
190   provider of the data, such as a Contract, Grant, Memorandum of Understanding (MOU), or
191   Data Sharing Agreement?
192   11.2.   If the source of the data is a NOAA Grantee, has the *Data Sharing by NOAA Grantees*
193   procedural directive been followed?
194   11.3.   If there are restrictions placed on NOAA use of the data, who will approve the use of these
195   data and assume any risk on behalf of NOAA associated with compliance with the terms of use?
196   11.4.   Has the NOAA Office of General Counsel cleared the agreement?
197

198 Appendix B. Related Policies

199 *NWS Policy Directive 1-12* and NWS *Instruction 1-1201*

200 • Focuses on rights and restrictions regarding data use and redistribution.

201 • http://www.nws.noaa.gov/directives/sym/pd00112001curr.pdf

202 *NOAA Web Mapping Applications Policy*

203 • States external data used in web mapping applications must be necessary for, and material to,
204 the presentation of agency information or the delivery of agency services; maps must credit or
205 provide a direct link back to the external provider; data must be relevant and timely, and
206 complete steps must have been taken to ensure that data layers are actively updated to achieve
207 the highest level of quality possible.

208 • http://www.cio.noaa.gov/Policy_Programs/NOAA_web_mapping_applications_policy.html

209 *NOAA Information Quality Act Guidelines*

210 • States that ensuring the quality of information is an important management objective that takes
211 its place alongside other NOAA objectives; defines key terms including Quality, Objectivity,
212 Reproducibility, Transparency, and Highly Influential Scientific Assessment.

213 • http://www.cio.noaa.gov/Policy_Programs/IQ_Guidelines_011812.html

214 *NOAA Procedure for Scientific Records Appraisal and Archive Approval*

215 • Describes how to decide what to archive and how to make arrangements with NOAA Data
216 Centers.

217 • https://geo-ide.noaa.gov/wiki/index.php?title=Category:NOAA_Procedural_Directives

218 *Data Sharing by NOAA Grantees* procedural directive

219 • States that grantees must include a data-sharing plan with proposals, and if funded must make
220 resulting data available in a timely fashion.

221 • https://geo-ide.noaa.gov/wiki/index.php?title=Category:NOAA_Procedural_Directives

222 *NAO 212-15: Management of Environmental data and Information*

223 • States that environmental data will be visible, accessible and independently understandable to
224 users, and will be managed based upon an end-to-end data lifecycle.

225 • http://www.corporateservices.noaa.gov/ames/administrative_orders/chapter_212/212-15.html

226 *NAO 212-13: NOAA Information Technology Security Policy*

227 • States that IT security will be applied throughout all phases of an information system's lifecycle.

228 • http://www.corporateservices.noaa.gov/ames/administrative_orders/chapter_212/212-13.html

229 Relevant US Government policies include:

230 • *OMB Circular A-130: Management of Federal Information Resources.*

231 • OSTP Policy: *Data Management for Global Change Research* (1991).