

# **Confidentiality and Data Quality Protocols for BSAI Crab Economic Data: A Discussion and Proposal**

by

The Economics and Social Sciences Research Program  
Alaska Fisheries Science Center  
Seattle, WA

Based on public testimony and a recommendation from the Advisory Panel at the December 2006 meeting, the Council passed a motion directing staff to develop protocols concerning data collected under the BSAI crab rationalization Economic Data Reporting (EDR) program. The protocols are to apply to two general areas, 1) maintaining data confidentiality and 2) assessing the quality of the data to ensure accuracy. To maintain confidentiality, the Council directed staff to develop protocols for Council review specifying aggregation requirements to avoid revealing proprietary data of fishery participants. This reiterates directions given in the February, 2003 Council Motion on Crab Rationalization: "The agencies will develop a protocol for the use of the data, including controls on access to the data, rules for aggregation of data for release to the public, penalties for release of confidential data, and penalties for unauthorized use." The December motion concerning data quality identified several areas of interest to the Council. First, the Council recommended that staff develop descriptions of data, their quality, deficiencies, and variability. These descriptions, in turn, would be used to draft protocols specifying appropriate scope of use for the data. In addition, descriptions would be used to determine appropriate revisions to the EDR questionnaires.

This discussion paper outlines the confidentiality and data quality issues to be resolved and the process that AFSC staff, in collaboration with Council and Region staff, will undertake to develop both sets of protocols to ensure that industry and Council concerns regarding the crab EDR program are addressed.

## **Confidentiality Protocols**

The BSAI crab EDR data contains detailed proprietary information provided by firms and individuals, as well as personally identifying information (PII) and business identifying information (BII). These data are considered confidential under the MSA and other federal statutes. NOAA Administrative Order (NOA) 216-100 is the principal legal guidance for NMFS employees on protocols for handling confidential data including definitions, policies, operational responsibilities and procedures, penalties, and statutory authorities. Before describing a plan to develop confidentiality protocols to apply specifically to crab EDR data pursuant to the Council motions, it seems appropriate to describe the baseline standards and procedures that apply generally to all confidential statistics collected and maintained by NMFS. The motions call for protocols to specifically address access to the data, penalties for unauthorized disclosure or unauthorized use of the data, and procedures for aggregation of data for use in any public release or reporting.

### Access

Access to the crab EDR data is tightly controlled under numerous provisions of statute, regulation and administrative order. The Code of Federal Regulations (50 CFR 600.405) specifies that access to confidential data is restricted to 1) federal and Council employees responsible for collection and maintenance of the data, FMP development,

monitoring or enforcement, or performing research that requires access to confidential statistics, or on a demonstrable need-to-know basis; 2) NOAA/NMFS contractors or grantees who require access to confidential statistics to perform functions authorized by a Federal contract or grant; or 3) state personnel who demonstrate a need for confidential statistics for use in fishery conservation and management, provided that the State has entered an agreement to protect confidential data to a standard comparable to that required by MSA. The regulations further provide for granting of access to Council members under conditions that are unlikely to be met in the case of the crab EDR data, and individual submitters may request that their own records be released to themselves or a third party. In addition, the confidential proprietary data collected in the crab EDR meets the definition of trade secrets as defined in the Freedom of Information Act (5 U.S.C. 552) and Trade Secrets Act (18 U.S.C. 1905), and as such is exempted from disclosure of raw, unaggregated data under FOIA. All individuals who are determined to be authorized for access to confidential data are required to sign and submit a nondisclosure agreement, affirming the user's understanding of NMFS obligations with respect to confidential data and the penalties for unauthorized use and disclosure.

NAO 216-100 establishes somewhat more detailed procedures for granting access to specific confidential data to NMFS employees, Councils and staff, state employees and contractors. Section 5.01 of the NOA confers responsibility for maintaining confidentiality of data collected within a given region to the Regional Director. 6.03b of the NAO specifies that "NMFS employees requesting confidential data must have certification as being authorized users for the particular type of data requested." Further, "authorized user" status may be granted under the NAO to Council members if approved by the Assistant Administrator, to state employees given approval of the NMFS office that maintains the source data, and to contractors if approved by the region. With the exception of Council members (for whom authorization authority is vested in the Assistant Administrator), the individual given authority to grant access to specific confidential data sources appears to be the Regional Director, although the language of the NAO is inconsistent on this point.

Further, pursuant to the Council motion authorizing the data collection, EDR data is collected by a third party, Pacific State Marine Fisheries Commission (PSMFC), which has custody of all original data. Prior to passing data on to NMFS, unique individual or vessel identifiers are screened by PSMFC. The only persons permitted to see identifiers are NMFS Alaska Region Restricted Access Management (RAM) personnel, NOAA General Counsel, NMFS Office of Enforcement, US Department of Justice, Federal Trade Commission, and PSMFC. To reduce the risk of identifier disclosures, with the exceptions above, no Council employee, NOAA employee, or State employee will have access the individual identifiers in raw EDR data.

*Proposed action*

Given the restrictions described above, we do not anticipate that any confidential information derived from the crab EDR program will be disclosed to the public or subject to unauthorized use, provided all individuals granted access comply with the law. However, given the sensitivity of these data, additional protocols for restricting authorization, at the discretion of the Regional or Science Center Director, to individuals working in specific capacities within or on contract to NMFS, ADF&G, or Council staff may be warranted. We propose to solicit input from agency leadership, the Council, industry and the public on further restrictions on access, and review procedures employed elsewhere within the Department of Commerce and other agencies. Based on

this input we will draft protocols for crab EDR data access authorization for consideration of NMFS leadership and Council consideration.

### Penalties

Persons who make unauthorized disclosure of confidential data, including confidential data collected in the crab EDRs, may be held to be subject to criminal prosecution under the Trade Secrets Act (18 U.S.C. 1905) and may be fined, imprisoned and removed from office or employment. Persons who unlawfully and willfully disclose agency records that contain individually identifiable information may be subject to criminal prosecution for violation of the Privacy Act (5 U.S.C. 552a(i)(1)) and may be fined. Department of Commerce employees are further prohibited by Department of Commerce employee conduct regulations and by ethics regulations applicable to the Executive Branch [5 CFR 2635.703] from using nonpublic information subject to this Order for personal gain, whether or not there is a disclosure to a third party. Apart from criminal prosecution, pursuant to NOAA Administrative Order 216-100 (relating to Protection of Confidential Fisheries Statistics) individuals may be subject to disciplinary action, including removal, for failure to comply with its provisions.

### *Proposed action*

As noted above, penalties for unauthorized disclosure of crab EDR data are quite strong. The potential for criminal prosecution, as well as disciplinary action under NAO 216-100 are expected to be sufficient disincentives to either intentional or unintentional disclosures, and we propose no additional penalties.

### Aggregation

While unauthorized disclosure of individual EDR records is clearly prohibited by the restrictions reviewed above, there is a need to present information derived from the data to the Council as well as the public. Federal law and regulations under which the EDR data are collected permit the release of information derived from confidential data that is structured to prevent identification of individual submitters or the information submitted by them. The procedures set forth in NOA 216-100 for structuring the data are not described in detail, however, and merely define aggregate or summary data as “data structured so that the identity of the submitter cannot be determined either from the present release of the data or in combination with other releases.”

A general rule of thumb that is applied for aggregation is that any unit (e.g. fishery, fleet, or sector) for which statistical information is reported must include at least three entities i.e., individuals, vessels, corporations, associations, or whatever form the data submitter takes). Under the data sharing agreement between ADF&G and NMFS, fisheries data supplied by ADF&G are to be aggregated over units of at least four entities. A single entity may own multiple vessels in the crab fishery and file as many EDRs. Therefore, to preserve the confidentiality of the EDR data in light of the complex ownership structure of the fishery, it may be necessary to aggregate over larger numbers of vessels in order to preserve confidentiality of all submitters' proprietary information. Alternately, other methods than simple aggregation may be used to preserve confidentiality. The Office of Management and Budget Committee on Statistical Methodology has reviewed the statistical methods employed by federal agencies charged with collecting extensive proprietary information. The committee has detailed their findings in the Report on Statistical Disclosure Limitation Methodology (OMB, 2005). Summarizing these results is beyond the scope of this discussion, but the report will be reviewed carefully for statistical methods that may be applicable to crab EDR data in the process of developing confidentiality protocols.

### *Proposed action*

A minimum standard for aggregation of the EDR data must satisfy criteria for both maintaining confidentiality as well as allowing for dissemination of information and analytical results at the resolution necessary to address relevant management concerns. It is clear that a simple standard of aggregating a minimum of three vessel EDRs is inadequate to protect confidentiality in light of the integrated ownership of the fishery. However, given available information, it is not possible to control perfectly for common ownership in defining an optimal aggregation. Therefore, the determination of the minimum standard for aggregation will most likely be a matter of reasonable precaution. We propose to solicit input from industry members who are subject to the EDR reporting requirement, as well as other interested individuals, regarding concerns about disclosure of social and economic information in aggregated EDR data. Based on public and industry input, review of the OMB report and consultation with other agencies including other NMFS regions, we will identify a range of alternative statistical procedures that address confidentiality concerns and present these for SSC review and review by the Council.

### **Data Quality Protocols**

The principal procedural requirements pertaining to quality of information disseminated by NOAA Fisheries are set forth under the federal Data Quality Act (Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001). NOAA Section 515 Information Quality Guidelines (NOAA, 2006) apply broadly to all information that the agency disseminates. Information that is collected for internal use by agency personnel and contractors, as is the case for the unaggregated EDR data, is not subject to the Data Quality Act requirements. However, any information that is synthesized from EDR data and subsequently disseminated by NOAA Fisheries, such as model results or aggregate-level statistics, is subject to the Act and covered by NOAA Information Quality Guidelines. As such, we consider NOAA Information Quality Guidelines as providing the relevant legal guidance regarding standards for data quality associated with the EDR program.

A key requirement of the Section 515 Guidelines is a “Pre-dissemination Review” that ensures the utility, integrity and objectivity of information released. These terms are defined in the Guidelines; in the context of this discussion paper, objectivity is of principal concern:

“Objectivity consists of two distinct elements: presentation and substance. The presentation element includes whether disseminated information is presented in an accurate, clear, complete, and unbiased manner and in a proper context. The substance element involves a focus on ensuring accurate, reliable, and unbiased information. In a scientific, financial, or statistical context, the original and supporting data shall be generated, and the analytic results shall be developed, using sound statistical and research methods.”

Pre-dissemination review standards are distinct for different types of information products. For the synthesized information that would be publicly reported from crab EDR data, the review standards are the following:

“Objectivity of synthesized products is achieved using data of known quality, applying sound analytical techniques, and reviewing the products or processes used to create them before dissemination.

Data and information sources are identified or made available upon request.

NOAA uses data of known quality or from sources acceptable to the relevant scientific and technical communities in order to ensure that synthesized products are valid, credible and useful.

Synthesized products are created using methods that are either published in standard methods manuals, documented in accessible formats by the disseminating office, or generally accepted by the relevant scientific and technical communities.

NOAA reviews synthesized products or the procedures used to create them (e.g. statistical procedures, models, or other analysis tools) to ensure their validity.

- Synthesized products that are unique or not produced regularly are reviewed individually by internal and/or external experts.
- For regular production of routine syntheses, the processes for developing these products are reviewed by internal and/or external experts.

NOAA includes the methods by which synthesized products are created when they are disseminated or makes the methods available upon request.”

The Guidelines recognize that where confidential data are concerned, the source data for synthesized products cannot generally be made available:

“Where confidentiality or other considerations preclude full transparency, then especially rigorous robustness checks will be applied. They may take many forms, ranging from the use of outside review panels to the use of an array of specific checks to ensure objectivity. The nature and a description of these checks will be disclosed upon request.”

Directions under the Information Quality Guidelines that apply particularly to the Crab EDR program are therefore to develop and apply “especially rigorous robustness checks.” A description of the process in place and to be further developed follows.

The EDR forms developed to elicit information on revenues, costs and other social and economic data from vessel operators are essentially measurement instruments, and like other scientific instruments are subject to some degree of measurement error. In the case of the crab EDRs, measurement error arises when information reported by an individual submitter in response to an elicitation for a given data element differs to some degree from that intended by the designers of the EDR forms. Given the complexity of the economic phenomena that take place within the harvest and processing sectors, it is impossible to design instruments that completely eliminate any error. There will most likely always be some degree of misinterpretation, and it may not always be possible for unique economic enterprises to characterize their operations in exactly the terms specified in a given data form. The objective for those tasked with designing and conducting a data collection program is to minimize error through careful design, and to characterize error, both quantitatively, in terms of variance and bias, as well as

qualitatively in the resulting dataset to permit proper interpretation by analysts. The relevant inquiry under the Information Quality Guidelines is whether the resulting dataset is within an acceptable degree of imprecision or error appropriate to the particular kind of information at issue (see NOAA Information Quality Guidelines, Part II, Objectivity).

Alaska Fisheries Science Center staff oversees the design and maintenance of the EDR data collection process, in conjunction with the third party data collection agent, Pacific States Marine Fisheries Commission (PSMFC). Procedures currently in place to monitor and improve data quality include the following:

1. Monitoring submitter feedback on EDR completion – Data collection staff at PSMFC work closely with individual submitters to clarify the intent of individual data elements and maintain a detailed log of questions and comments that submitters provide through written and verbal communication. The feedback documented in these logs is used to identify any consistent pattern of misinterpretation and the potential for misreporting.
2. Data audit - A component of the EDR program specified in authorizing legislation is the compulsory audit of EDR forms to identify intentional and unintentional misreporting. During summer and fall of 2006, the protocols for implementing both random and outlier audits of EDRs submitted for the years 1998, 2001, 2004, and 2005 were developed and a professional accounting firm (Aldrich, Kilbride & Tatone, Portland, OR) was hired to complete the audits. A report detailing the results of the analysis and providing qualitative and, where possible, quantitative error measurements of selected data elements, is due from the auditors in March, 2006. The results of this analysis will be used to both improve the EDR forms to reduce error as well as to characterize the quality of the data collected thus far to guide analysts in proper use and interpretation of the data.

#### *Proposed action*

A briefing paper summarizing the results of the audit and the feedback from EDR submitters will be developed. Public meetings with industry members and other interested individuals will be scheduled for summer, 2007 to review the report and solicit further input. Meetings for harvesters, processors, and the general public will be held. Possible meeting locations include Seattle, Kodiak, and Anchorage (meeting announcements will be published in the Federal Register), pending input from the Council. These qualitative and quantitative results will be used to revise and improve the EDR forms to reduce error in data collected in the future (see below for proposed timeline). The results will also be incorporated into a metadata document that will accompany the dataset, describing sources and degree of error on a variable-by-variable basis. Variables which have been indicated through submitter feedback and the audit process to have large error rates will be identified and analysts cautioned against misinterpretation or unwarranted use. Deletion of individual variables or other components of the database from distribution to authorized users will be considered, subject to legal requirements for database maintenance. The metadata document (which will not contain actual EDR data in aggregate or disaggregate form) will be available for public and Council review.

We propose to establish specific protocols for formal analysis and reporting of the crab data necessary to preserve confidentiality and require reporting of uncertainty of analytical results. In addition, appropriate statistical and analytical methods that incorporate data quality will be achieved by restricting authorization for access to the data to individuals credentialed to use the data, and through internal peer review and

SSC review.

The EDR data will not be used by ASFC staff, contractors, or other authorized individuals prior to completion of the metadata document and confidentiality protocols. Prior to release of any analytical or information products from EDR data, including aggregated data and model results, the release will undergo Pre-dissemination Review as required by NOAA Section 515 Guidelines.

### **Proposed timeline**

March - June, 2007

- Receive Audit report

- Prepare briefing paper on confidentiality and data quality protocols and procedures

July, 2007

- Hold public meetings in Anchorage, Kodiak, and Seattle for industry and other members of the public to review briefing paper and solicit further input.

August - September, 2007

- Incorporate public input and develop draft protocols for confidentiality and data quality. Circulate draft for internal peer review.

October, 2007

- Submit preliminary metadata document and proposed data quality and confidentiality protocols for Council and public review and comment. Additional actions as necessary.

### **References**

National Oceanic and Atmospheric Administration. 1994. NOAA Administrative Order 216-100. : Protection Of Confidential Fisheries Statistics.

[http://www.corporateservices.noaa.gov/~ames/NAOs/Chap\\_216/naos\\_216\\_100.html](http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_100.html)

National Oceanic and Atmospheric Administration. 2006. Information Quality Guidelines November 6, 2006 Revision.

[http://www.cio.noaa.gov/itmanagement/IQ\\_Guidelines\\_110606.htm](http://www.cio.noaa.gov/itmanagement/IQ_Guidelines_110606.htm)

Office of Management and Budget. 2005. Federal Committee on Statistical Methodology, Statistical Working Paper 22: Report on Statistical Disclosure Limitation Methodology. 137 p.