2 Executive Summary

An executive summary of the defining characteristics and main aspects of the "ideal system" as defined in this document follows. The rest of the document expands on the schematic in the previous section and these brief summaries:

Data-Centric: The design of the ideal system is data-centric. The central asset of the organisation is a central database holding all data. Supplier data is ingested into it, transformed into a product neutral representation and output into multiple products as required by end user stakeholders.

Data-Centric: The rest of the infrastructure, support, services and, to some extent, organisational structure is driven by the data in the central database.

Product-Neutral: The ideal system is product neutral – the internal representation of the real world features is defined and driven by an overarching data model and model-driven architecture principles define physical database characteristics, interfaces, application schemas, formats and all technical artefacts required for the system architecture. This is independent of the COTS software making up the system.

Data Ingest: Ingest into the ideal system is facilitated by international agreements with data suppliers for pre-compiled products and other sources. Trusted data suppliers feed data which is pre-validated and, ultimately, is in a model-ready form for seamless ingest into the database. Data is horizontally and vertically consistent to facilitate the creation of the seamless, product neutral database within NGA. A comprehensive metadata structure is built around ingested data to track its source, provenance and usage within multiple products on extraction.

Technology Development: What the Agency will achieve with a gridded, derived, annotated and edited global ENC (+AML +SMENC) series is novel and technically innovative at a fundamental level to what other providers have today and will require tooling which is unavailable in most COTS software. There are several ways to achieve the ingest, processing and extraction but to avoid large amounts of rework and manually intensive updating and re-scheming a set of tool extensions and feature-level processing tools will be required which blend geospatial knowledge with committer-level competencies in current and future ENC technology and standards. There are probably enough of these components to warrant the a common architecture and platform on which to host and integrate them with the main system.

Data Transformation: On ingest, data is transformed into a reconciled, product neutral, seamless form ready for export.

Data Editing and management: As necessary, the data in the central database can be enhanced, annotated and edited according to NGAs network of data suppliers. Although some data (the "40%") is trusted and runs through the system without edit, much of the data is validated, edited and prepared for output. Other data suppliers' data is conflated with NGA sources, assessed and documented.

Multiple Product extraction: The product neutral database model allows for the extraction of multiple end user products from the feature instances held within it. The model references existing standards under the ISO191xx framework and uses IHO S-100 structures to facilitate output of data in an intermediate format. This format can be transformed into multiple standards conformant products ready for dissemination to end users.

The Grid: Extracted products all align to a single, logically defined grid scheme which balances NGA requirements, brokered and non-brokered data content and which works at a practical level for current and future products.

Updates: The system can produce products either by reissue (all features disseminated to users) or by incremental update of individual features as required.

Generalisation: At extraction, scaleless layers in the central database and automated generalisation processes generate content within a banded range of scales ready for transformation into end user products.

Extraction: At extraction features are clipped intelligently to a pre-defined grid structure upon which all end user products are defined. The grid system imposes common boundaries between scales and has an intuitive naming scheme to assist end users. The grid has an origin and is optimised for ingest of supplier data and its efficient compilation into end products for primary navigation. The grid definition and its external nature reinforce the product-neutral nature of the system solution.

3 Introduction and Background.

The purpose of this report, PWS3, is to detail the "ideal future state" of NGA Maritime production of data for both primary navigation and broader use. This report contains a detailed description of that future state, the "ideal system" and its defining characteristics. In line with the data-centric nature of NGA Maritime's strategic direction this report is, itself, presented as a data-centric methodology in that we focus primarily on the idealised future data assets and their characteristics, building a surrounding infrastructure around them and describing how they interact with the central data assets.

This report forms a picture of the "ideal" solution for NGA Maritime long term and is a prelude to a more practical look at processes, technology and tools which might be used to achieve it (the subject of future work items for the project).

Ancillary sections detail the international environmental assessment current at the time of writing and documents the current relevant standards development efforts, likely timeframes and initiatives which may be of interest and which are relevant as well as some other specific topics of interest – a previous paper written on the NGA proposed naming scheme for ENCs is also included for completeness at Annex A

Once the ideal system is described fully, individual parts which best exemplify the approach will be listed as candidates for testing in a real world context later in the project (PWS4). The aim of the next phase is to test the areas of functionality and technical complexity which are most challenging or key to the success of the operation of the idealised system in the future

3.1 Overview – Problem Statement

What we are trying to achieve with the project is the construction of a view of the core NGA database which facilitates the following activities:

1. The migration of the NGA maritime data production system to one which is solely focused around the promulgation of ENC, AML and SMENC data from multiple sources, both brokered non-brokered and partially or wholly compiled from other source material.
2. The adoption of a data-centric approach. This focuses on definition of core data assets and then builds infrastructure around them.
3. A more product neutral approach to the storage of data assets and their extraction into the end user community
4. The implementation of a gridded scheme globally, for all products, and the details of that grid's characteristics e.g. scale, size, currency etc.

These were explored in detail in the initial view of the NGA systems observed by the team in the first site visit and are used as strategic principles for the content of this report.

3.2 Overview - the "Database View"

The schematic diagram (shown in Figure 1) provides the main guiding source for this report which shows how the system is broken down into functional elements which ingest, facilitate, process, store and output data. The "ideal system" defined in this document is one where a large quantity of data

from brokered and non-brokered relationships is deconflicted and transformed into a product neutral model (separately developed) and then stored, managed and processed within a single database structure. The data within it is able to be extracted into multiple products for use by NGA key customers as well as broader MSDI usage.

The system implements a suite of automated processes on extract of product features which can generate generalisation as well as the implementation of scaleless layers ensuring vertical consistency of features between scale bands and minimising any manual compilation effort.

The broad picture of the "ideal" solution documented here is broken into a set of four major components. These represent the core conceptual elements of the ideal system:

1. Data input, International Relations and Bilateral and agreements. Section 0
2. Data Ingest, Storage, Management and Processing. Section 5
3. Model Driven Enterprise Architecture. Section 9
4. Outputs, Validation, Delivery and Broader Usage. Section 10

These four major components are subdivided in individual sections later in this document and supplemented with details of their implementation. An additional major section, Section 8, looks in detail at the grid definition under development and proposes some alternative and complementary approaches.

The executive summary at the start of this document contains a (very) short description of how the "ideal" system is intended to function.

3.3 Technology Development.

As well as the four elements of the ideal system defined in the previous section it is worth stating early on that the process of construction of a truly product neutral database will require a considerable engineering effort and is certainly, at least to our knowledge, not available in a COTS form today. This required technology will be focused in two main areas which NGA will be pioneering at this scale:

1. The need to be truly product neutral can be equated with the ability to perform automated generalisation of all feature classes at a full range of scales and to create a model of product neutral features capable of simultaneous extraction into multiple user products.
2. The ability to "derive" ENC charts (both S-57 and S-101) from brokered and non-brokered data including their incremental update, tracking of persistent unique identifiers and conflation of in-house editing all within a highly regulated and standardised validation process.

These are not insurmountable tasks by any means and all the tools exist to assemble such a system but the construction of such a facility from COTS will require engineering in the above areas to make it truly product neutral and to automate processes as far as possible. This should be viewed as a continuum though and a process of migration and development of key technologies will be necessary in the early phases to de-risk the enterprise. Accordingly the next phase of the project defines some of the key activities where technical risk exists.

This report does not concentrate on the technical details of how these issues may be solved. They certainly require further research and are also dependent on both COTS and bespoke technology (and their integration) to realise the possible benefits of more automation. Elements of these technologies and key algorithms will be proposed for development in the next phase of the project.