International Association of Geodesy Symposia

A standard file format for gridded geodetic data

--Manuscript Draft--

Manuscript Number:	
Full Title:	A standard file format for gridded geodetic data
Article Type:	Original Article
Corresponding Author:	Kevin M. Kelly, M.A.Sc. Esri Redlands, CA UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Esri
Corresponding Author's Secondary Institution:	
First Author:	Kevin M. Kelly, M.A.Sc.
First Author Secondary Information:	
Order of Authors:	Kevin M. Kelly, M.A.Sc.
	Rob Juergens, B.A.
	Melita Kennedy, M.S.
	David Burrows, M.S.
	Roger Lott, B.Sc., FRICS
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	The use of gridded data in geodetic applications is increasing. For some geodetic datum transformations the Natural Resources Canada NTv2 format is widely used although other national and ad hoc formats are also in use. There is no global standard for these and other geodetic data such as geoid grids, position displacement grids and numerous others. Producers often define a proprietary or some other convenient format. The adoption of a standard file format will facilitate the creation and use of gridded data sets. It would relieve grid producers of the necessity for producing file readers and will assist application developers to incorporate new grids with minimal effort. Users will benefit from quicker access to the data. This paper describes the business case for and technical requirements of an international standard for a geodetic data grid exchange format (GGXF). It outlines a proposal for producing an international standard.
Suggested Reviewers:	

A standard file format for gridded geodetic data

Kevin M. Kelly, Rob Juergens, Melita Kennedy, David Burrows
Environmental Systems Research Institute (Esri), Redlands, CA, USA

Roger Lott
International Association of Oil and Gas Producers (IOGP), London, UK

Abstract. The use of gridded data in geodetic applications is increasing. For some geodetic datum transformations the Natural Resources Canada NTv2 format is widely used although other national and ad hoc formats are also in use. There is no global standard for these and other geodetic data such as geoid grids, position displacement grids and numerous others. Producers often define a proprietary or some other convenient format. The adoption of a standard file format will facilitate the creation and use of gridded data sets. It would relieve grid producers of the necessity for producing file readers and will assist application developers to incorporate new grids with minimal effort. Users will benefit from quicker access to the data. This paper describes the business case for and technical requirements of an international standard for a geodetic data grid exchange format (GGXF). It outlines a proposal for producing an international standard.

Keywords: *grid format, geodetic grids, standards, ISO, EPSG, GIS*

1 Why do we need a standard geodetic grid file format?

Grid files have been used in geodetic applications for over a quarter of a century. The U.S. National Geodetic Survey's (NGS) NADCON (*Dewhurst*, *1990*) grids of latitude and longitude differences were introduced before 1990, with Natural Resources Canada's (NrCan) National Transformation (NT) (*NrCan*, *2014*) format following in Canada not long afterwards. Other countries have followed similar approaches, for example the Ordnance Survey National Grid Transformation (OSTN) and grid (*OSGB*, *2015*) in Great Britain and the GR3DF97A transformation and grid (*IGN*, *1997*) in France. These four approaches utilize grid files with different content and different format. Led initially by Australia and followed by several others, there has been further adoption of the Canadian NTv2 format which has become a de facto standard for 2-dimensional transformations using latitude and longitude difference files.

For geoid and height correction models there has been no similar global adoption of one format. Almost every producer of these models has defined their own file format. For example the US NOAA/NGS Vertical Datum Transformation (VDatum) (*NOAA*, *2012*) and UK Vertical Offshore Reference Frame (VORF) (*UCL*, *2012*) applications for marrying hydrographic sounding datums, gravity-related (orthometric / normal) heights and ellipsoid heights take different gridding approaches, as do the U.S. National Geospatial Intelligence Agency's (NGA) Earth Gravitational Models (*NGA, 2013*) geoid files of various epochs (e.g. EGM84, EGM96 and EGM2008) (*NGA, 2013*; *NGA, 2014*) and the French approach to geoid files for both metropolitan France and French overseas departments and territories (*IGN, 2012*). The situation is summarised by *Neacsu, 2010* as:

> "Geoid files are presumed to be simple stuff as they contain just one value (undulation – the separation between ellipsoid and geoid) in a regular, rectangular grid pattern. You would have expected that the few hundred people who are involved in generating those models would be able to agree on a format and follow that standard. Alas, the reality is far different. For us the story began at some point in the '90s when we implemented the GEOID96 model for continental US. Figuring that there wasn't much merit in inventing a new file format, we implemented the USGS format at the time: a binary format with the extension .GEO. Not long after, USGS decided to change the format and came out with another one called BIN. It didn't take long for other organizations to come out with new formats, some ASCII some binary, some going from North to South, others going from South to North, from East to West and from West to East. You get the picture.
>
> Luckily, we decided to stick to one format and created just a few conversion programs that can read and convert all the other formats and produce the GEO file. The rest of the article details the different formats we can currently read and convert. If you need a new geoid model integrated in [our product], we can do it a lot easier if it is in one of these formats. However, we understand that in most cases the format of the data is not something that you, our user, can control but something created by a big bureaucratic organization and we will do our best to try to work with it. Just don't expect that we can do it overnight."

This has led to the following limitations.

- Grid producers have often had to supply some software to read and interpolate their gridded data.
- Each vendor of globally-marketed applications has had to repeatedly write programs to import the different grids. This has often resulted in reformatting of the grid file content to each vendor's proprietary internal format.
- Users of globally-marketed applications have been unable to use gridded data until it has been incorporated into the application. At best this causes a time delay, sometimes years. At worst there may be no economic benefit to the application vendor to incorporate highly specialist grids and the client user company is unable to access the gridded data. Global corporations such as international oil companies usually have strict procedures for the acquisition of new software, causing difficulty in adopting grid-producers' interpolation software. If such corporations are to use the gridded data, it needs to be through one of their internally-authorised applications.

Were there to be a standard grid file format then:

- Grid producers would not have to provide their own software for each data format.
- Survey and GIS software vendors would need to read only one grid file format.
- End users could use a new grid file as soon as it became available, without having to wait for their application vendor to produce a software upgrade.

The benefit of a standard file for geoid models and height correction models is easily demonstrated. The forthcoming convention for defining gravity-related height reference surfaces is through offsets from a three-dimensional Earth reference model. More and more such models are being produced and this trend looks likely to continue.

But from a computing perspective, the mechanisms for reading and processing data from a gridded data set are largely uncorrelated with the intrinsic meaning of that data. It therefore makes sense to seek a solution that would be general to many different geodetic data types: horizontal datum transformations; vertical offsets for a wide variety of applications: geoid models, sea surface topography, glacial isostatic adjustment (GIA), etc; velocity models and many others. A generic geodetic data grid exchange format would then meet a wide variety of potential applications.

The remainder of this paper discusses requirements for and challenges of defining a software independent standard Geodetic Grid eXchange Format (GGXF).

2 Features required of a standard geodetic data grid format file

We have identified the following requirements of a GGXF file.

Informative header. Most existing grid format headers provide no information about what the gridded data represents. One must rely on external documentation for this description and, if not available, the grid is rendered meaningless. This malady plagues not only geodetic grid headers but also others as well (e.g. Esri ASCII raster format). The proposed format must allow for a full description of the data contained in the grid, including but not limited to coordinate reference systems and data provenance.

User definable content. The proposed GGXF format will allow users to define the type of data represented by the grid. A predefined list of well-known geodetic data types will be part of the format specification. Users should be able to propose additions to this list which can readily be added to the specification.

Multi-dimensional. Most existing geodetic grid formats permit only pre-defined data values at each grid node. The proposed GGXF format will allow users to define a set or tuple of numeric data at each grid node, $(d_1, d_2, \ldots, d_i)$, where d_i is the *ith* data value in the tuple. For example, three-dimensional velocity and accuracy data could be defined at each grid node as the sequential set $\left(v_X, v_Y, v_Z, \sigma_X, \sigma_Y, \sigma_Z\right)$, in an Earth-Centered, Earth-Fixed Cartesian reference frame such as ITRF2008 or, equivalently in ellipsoidal geodetic coordinates as $\left(v_\varphi, v_\lambda, v_h, \sigma_\varphi, \sigma_\lambda, \sigma_h\right)$, where φ, λ, h are geodetic latitude, geodetic longitude and ellipsoidal height, respectively.

Multi-resolution. The proposed format must cope with multiple levels of data resolution.

Interpolation. The file header should allow for the specification of the required or recommended interpolation method, e.g. bilinear, spline, etc. This will allow application software to apply the grid producers' desired interpolation algorithm.

Efficient. GGXF is required to engender computational efficiency through supporting binary files, platform independence and efficient file size. Grid header location will be flexible: headers for all grids can be listed sequentially at the beginning of the file or headers can precede each grid as it appears in the file. Since, parent grids and subgrids can appear in any order in the file, the flexible header facilitates fast access to any desired part of the file. With these, developers can build applications to have fast access to any part of very large files.

Amongst our other design goals are to ensure linguistic and culture adaptability and ease of adoption.

Human readable. ASCII file support will allow human readable file content and aid users preferring to work with ASCII file types.

Easy to use. The proposed format should be straightforward for grid producers to populate and easy and efficient for application developers to use. A well-documented description, complete with examples and an open-source API for developers must accompany the GGXF format. GGXF standard file syntax will be published and versionable.

Recognizable. GGXF file format will be identified by file name extension: [*filename*].GXA for ASCII file types and [*filename*].GXB for binary files.

3 Existing formats

We mentioned earlier a sampling of geodetic agencies that produce gridded geodetic data products, i.e. NGS, NrCan, NGA and others. Within these agencies we find that different gridded products use different formats. Consequently, neither inter-agency nor intra-agency consistency prevails for these geodetic products. For example, NGS has published such geodetic grids as NADCON and HPGN grids for horizontal datum transformations, VERTCON and VDATUM for vertical and tidal datum transformations and numerous geoid model grids, few of which share a common data format and they thus require separate specialized readers for each. A similar situation prevails for many of the other agency formats as well.

Some of these aforementioned formats, for example NGS's NADCON, HPGN, and GEOCON, use multiple files for the shift in different coordinate directions rather than a single file holding all shifts. A single file would yield more efficient and faster access of the data files. Moreover, none of the existing formats have sufficient descriptive information to identify the data, its intended use and source, or even its file format description. This means that if a user obtains a copy of only the file, it cannot be used without additional external information.

In contrast to the geodetic community, the scientific geophysical community use altogether different grid formats such as Network Common Data Form (netCDF) and Hierarchical Data Format (HDF).

NetCDF is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. It is also a community standard for sharing scientific data (*UCAR, 2015*). NetCDF has grown into a *de facto* standard within for example the atmospheric modelling, meteorological and oceanographic communities (*UCAR, 2015*).

HDF is a data model, library, and file format for storing and managing data. It supports an unlimited variety of data types, and is designed for flexible and efficient input and output (I/O) and for high volume and complex data. HDF is portable and extensible, allowing applications to evolve in their use of HDF. The HDF Technology suite includes tools and applications for managing, manipulating, viewing, and analyzing data in the HDF format. HDF allows hierarchical data objects to be expressed in a very natural manner, in contrast to the tables of relational database. Whereas relational databases support tables, HDF supports n-dimensional datasets and each element in the dataset may itself be a complex object. Relational databases offer excellent support for queries based on field matching, but are not well-suited for sequentially processing all records in the database or for subsetting the data based on coordinate-style lookup (*The HDF Group, 2014*).

There are some advantages to adapting netCDF to the proposed geodetic grid format. NetCDF satisfies all the required features of GGXF, is well-established in the scientific community, is platform agonistic, has existing application interfaces (API), is actively developed and maintained and has over 100 open-source and commercial software packages that may be used for manipulating or displaying netCDF data (*UCAR, 2015*). These are compelling features that warrant thoughtful consideration; they serve as a deterrent to designing a new grid format. However, the broad flexibility of netCDF engenders a complexity that seems unnecessary to the rather simple nature of gridded geodetic data. This is not to say that netCDF is not a good option for GGXF, but only that further evaluation is needed to assess how well and how simply it can be adapted to gridded geodetic data. Adoption of netCDF for GGXF certainly minimises re-inventing data structures that already exist and are well-understood. It is noteworthy that NetCDF and HDF are rarely used by national geodetic agencies or academic geodesists, but this fact ought not to hinder a serious look at netCDF for GGXF.

4 Challenges

The format will only be judged successful if it becomes the standard, thereby removing the problems of the multitude of existing formats. By far the greatest challenge has been to marry the desire for the format to be both near self-defining and extensible and easy to use. An outline of the proposed file format and how the chosen attributes address these requirements is discussed below.

4.1 File structure

The GGXF file will consist of a header followed by the data. The header will use keywords and formatted fields to specify the parameters of the grid. The header need not share the same datatype as the data portion of the file. For example, the data could be binary and the header in XML, or binary data and an ASCII plain text header. The decision on header datatype will depend on considerations such as ease of implementation, availability of existing software to read and write GGXF headers, familiarity with the datatype, etc. We want a file structure that is not an impediment to the format's acceptance and makes for almost effortlessness adoption by users.

The requirement for differing levels of resolution suggests the use of sub-grids for detailed areas, as has been done for example in the NTv2 format. NTv2 encodes a header before each sub-grid; grids paired with their corresponding headers can be in any order in the file. An alternative approach which offers faster access uses a stacked header approach. In the stacked header method, a main header precedes multiple sub-headers, one for each grid, and listed prior to the start of the data stream. These successive headers can be in any order. A design issue to be resolved is which of these approaches GGXF should follow, or whether it should allow both. The stacked and in-file header approaches are illustrated in figure 2.

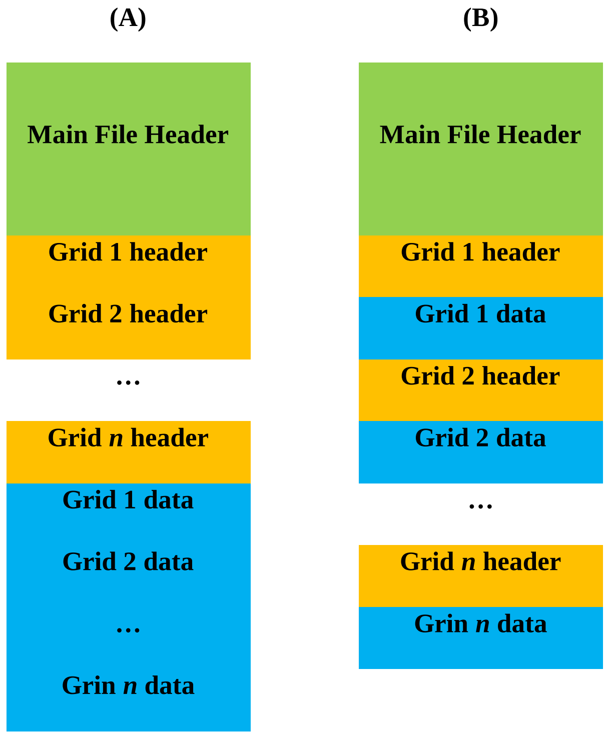

Figure 2. GGXF stacked header structure followed by data grids. (A) stacked header file method and (B) in-file header file method.

Essentially a GGXF file will comprise parent grids and child grids, or sub-grids. GGXF sub-grids will follow a similar structure as NTv2. To ensure fast access to large files, all grids must be rectangular. Each child grid must be bounded by an integer multiple of parent grid values. For example, if nodes in a parent grid were at say 5 degrees, then a child grid could start at 2.5 or 1.25 degrees, but could not start at 3 degrees. Figure 3 illustrates the structure of a GGXF parent grid and child grid.

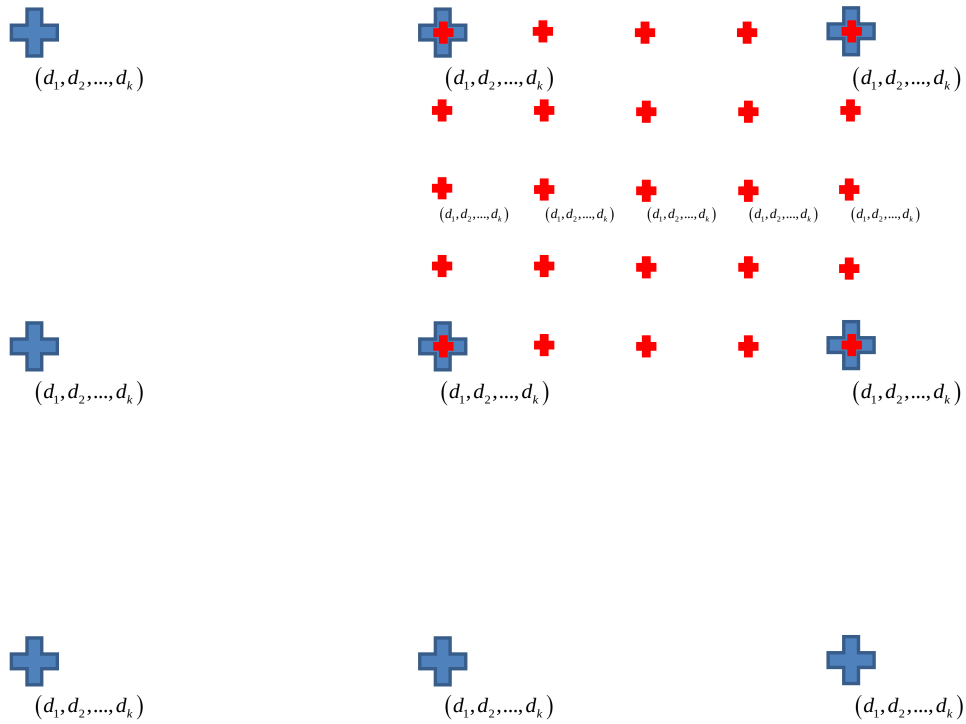

Figure 3. Structure of parent grid (blue crosses) and subgrid (red crosses) with data tuple shown in parentheses at data nodes.

The rectangular grid and sub-grid structure is required for fast access. Some datasets may have no data over parts of the rectangle. Null data flags will be allowed.

As previously mentioned GGXF is designed specifically for geodetic data. For that reason, and to minimize its complexity, GGXF version 1 requires grid node coordinates in a so-called 2D Geographic or 2D Projected horizontal coordinate reference system (CRS). A 2D Geographic CRS is a two-dimensional geodetic system with orthogonal ellipsoidal axes of latitude and longitude (*IOGP*, *2009*). A 2D Projected CRS is a two-dimensional projected system with orthogonal planar axes of *X* and *Y* or *Easting* and *Northing*. We envision future versions of GGXF which can accommodate three-dimensional CRSs, store three-dimensional gridded data and manage temporal data. Future versions will be backwards compatible.

Coordinates belong to a coordinate system. A Coordinate System (CS) describes the mathematical rules governing the coordinate space including: the number of axes, their name, their direction, their units, and their order. When coordinates are used to describe position on the earth, they belong to a Coordinate Reference System. A coordinate reference system (CRS) is a coordinate system which is referenced to the earth (*IOGP, 2008*). The particular CRS of a GGXF file must be defined in the main header and must be the same for all grids in the file.

CRS's are defined in the GGXF main header and may also include references to external registries to allow authentication. These registries may come from various authoritative agencies, and include: (1) EPSG[1], (2) ISO[2] and (3) Esri[3]. The EPSG database of geodetic information includes coordinate reference systems and their component parameters. It is maintained by the *IOGP Geodesy Subcommittee*. It can be accessed at no charge through an online registry at http://www.epsg-registry.org or downloaded as a database or SQL files from the IOGP EPSG website at http://www.epsg.org. The geographic coverage of the EPSG dataset is worldwide, but does not and cannot record all possible geodetic parameters in use around the world.

Though not yet available, the *ISO Registry of Geodetic Codes and Parameters* is a structured database of coordinate reference systems and transformations of international extent that will be accessible online. The registry will be provided under the auspices of ISO Technical Committee 211 on geographic information/geomatics and conforms to ISO standards ISO 19111:2007 (Spatial referencing by coordinates), ISO 19135:2005 (Procedures for item registration) and other supporting ISO standards (*Craymer*, *2014*).

Esri, makers of the ArcGIS® line of GIS software products, also maintains a system of codes for coordinate reference systems and transformations. This list incorporates the EPSG codes but supplements them with unique codes for CRS's not included in the EPSG set. The EPSG/Esri CRS codes form a *de facto* standard within a broad sector of the geomatics community. But to cater for a case where the CRS or frame is not defined in any publicly-accessible registry the GGXF format must also allow for specification of a user-defined system. The approach taken will be similar to the Well Known Text (WKT) representation of coordinate reference systems (ISO 19162:2015).

4.2 Ease of Adoption

To facilitate adoption we believe that the following is essential:

a) The file definition is clearly documented and easily available.
b) Conversion software from popular existing formats to GGXF is made freely available.
c) Tools for testing compliance with the format are available.
d) The format is well publicised in geodetic and GIS circles.

We intend to provide documentation in the English language, a free, open-source API and sample code to work with the format. Sample main header files for common geodetic grids will also be distributed.

---

[1] EPSG Geodetic Parameter Dataset.
[2] International Standards Organization. http://www.iso.org/iso/home.html
[3] Esri. http://www.esri.com/

5 Future plans

Receiver Independent Exchange Format (RINEX) is a data interchange format for raw satellite navigation system data (*IGS, 2013*). RINEX is one of a number of well-known formats in the geodetic community. It originated in the University of Berne as a mechanism for combining data from a wide range of different GPS receivers used in the initial EUREF campaigns. It has been a *de facto* standard for over a quarter of a century. Its documentation was maintained at the University of Berne although more recently that has been transferred to an International GNSS Service (IGS) working group in conjunction with the Radio Technical Commission for Maritime Services (RTCM). Its success has spawned similar formats, for example the exchange format for phase centre variations of geodetic GNSS antennae (ANTEX). The format has been successful without being a *de jure* ISO standard.

Initially we propose that the GGXF specification will be maintained by Esri and IOGP, with documentation available from both organisations. However a number of government agencies have made it clear to us that they would like there to be an ISO standard for gridded geodetic data. They argue that this will help minimise the proliferation of different formats currently seen. We see merit in these arguments and therefore are proposing to submit a draft standard to ISO TC211, either directly or possibly through the Open Geospatial Consortium (OGC) who have a strong application developer community but with formal links to ISO TC211.

6 Conclusions

No global standard file format exists for gridded geodetic data, yet these data are widely used within the geomatics and geophysics communities. Data producers often define a proprietary or some other convenient format. Data consumers are often burdened with coding reading software for each unique format. Some grid files are devoid of any in-file descriptive information about the meaning and use of the stored data. A standard file format for gridded geodetic data as we have proposed here will remedy these and other problems. It will:

- release grid producers from having to provide their own software for each data format.
- release survey and GIS software vendors from having to write reading software for numerous proprietary geodetic grid file formats.
- allow end users access to a new grid file without having to wait for their application vendor to produce a software upgrade.

We have outlined the general file structure and some features of our proposed **G**eodetic **G**rid E**X**change **F**ormat or GGXF. GGXF is still a work in progress and we expect to have a prototype GGXF specification complete with code and documentation by early to mid 2016. Initially we propose that the GGXF specification be maintained by Esri and IOGP, we see its potential as an eventual ISO standard.

References

Dewhurst, WT. NADCON - The Application of Minimum-Curvature-Derived Surfaces in the Transformation of Positional Data From the North American Datum of 1927 to the North American Datum of 1983. NOAA TM NOS NGS 50, 1990.

Institut Géografique National (IGN). “Grille de paramèters de transformation de coordonnées GR3DF97A”. April 1997. http://geodesie.ign.fr/contenu/fichiers/documentation/algorithmes/notice/NTG_88.pdf

Institut Géografique National (IGN). “Grilles de conversion altimetric”. July 2012. http://geodesie.ign.fr/index.php?page=grilles

International GNSS Sevice (IGS). “RINEX: The Receiver Independent Exchange Format”. April 2013. http://igscb.jpl.nasa.gov/igscb/data/format/rinex302.pdf

International Association of Oil & Gas Producers (IOGP). Surveying & Positioning Guidance note 5, Coordinate reference system definition — recommended practice. Version 2.1, January 2009.

International Association of Oil & Gas Producers (IOGP). Surveying & Positioning Guidance note 1, Geodetic awareness guidance note. Version 1.1, August 22, 2008.

International Standards Organization (ISO). ISO/IEC 13249-3:2011 Information technology -- Database languages -- SQL multimedia and application packages -- Part 3: Spatial. August 22, 2011.

National Geospatial-Intelligence Agency (NGA). “NGA/NASA EGM96 Earth Gravitational Model”, November 18, 2014. http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm96/egm96.html

National Geospatial-Intelligence Agency (NGA). “Earth Gravitational Model 2008 (EGM2008)”, May 06, 2013. http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/index.html

Natural Resources Canada (NrCan), “Tools and Applications”, May 27, 2014. http://www.nrcan.gc.ca/earth-sciences/geomatics/geodetic-reference-systems/tools-applications/10925

Neacsu, Mircea “Information about Geoid File Formats”, Hypack Inc. Newsletter http://www.hypack.com/new/portals/1/PDF/sb/07_10/New%20Geoid%20Model%20for%20France%20and%20Information%20about%20Geoid%20File%20Formats.pdf

NOAA. (2012). “Vertical Datum Transformation, Integrating America's Elevation Data”, National Geodetic Survey (NGS), Office of Coast Survey (OCS), and Center for Operational Oceanographic Products and Services (CO-OPS). http://vdatum.noaa.gov/welcome.html

Ordnance Survey of Great Britain (OSGB), “National Grid Transformation OSTN02”, 2015. http://www.ordnancesurvey.co.uk/business-and-government/help-and-support/navigation-technology/os-net/surveying.html

University Corporation for Atmospheric Research (UCAR), Unidata Program Center, Network Common Data Form (NetCDF), 2015. http://www.unidata.ucar.edu/software/netcdf/#home

University College London (UCL). “Vertical Offshore Reference Frames”, May 26, 2012.
http://www.ucl.ac.uk/vorf