

Project no. 30956

Project acronym: GNU

Project Title: GMES Network of Users

Instrument: Coordination Action

Thematic Priority: Aeronautics and Space

Deliverable Number: 2.4b

Criteria for an optimised setup of GMES and related Data portals

Due date of deliverable: Month 30

Actual submission date: Month 31

Start date of project: October 2007, Duration: 3 years

The work for this deliverable was co-led by
Environment Agency Austria (UBA-A) and European Forestry Institute (EFI)

Project co-funded by the European Commission within the Sixth Framework Programme
(2002-2006)

Content

Working method and purpose of this document.....	3
The GNU partner's acquaintance with data systems	4
Strengths and weaknesses of different data portals	6
Advantages of obtaining data-products through portals.....	9
Incentives and barriers to register products in a data system.....	11
Essential meta-data	13
Integrated user feedback.....	15
Portal authenticity	16
Beyond portals and catalogues.....	17

Working method and purpose of this document

Currently, GMES and related information is distributed via a wealth of different portals, partly on project, partly on overarching levels. Such portals give access to services or value-added products or raw data, and they show a large variety in terms of user friendliness, transparency, handling, access, graphics, etc. The way these portals are designed may either be an incentive or an obstacle to incorporate the offered data products or services in daily work practices and is, therefore, decisive as to the user uptake of GMES results.

The detailed planning for GNU's work on data portals was presented and discussed during a half-day plenary session at the fourth GNU Consortium meeting (Berlin, Germany, 8-10 June 2009). The Consortium members had been asked to prepare for that meeting by working through a set of seed-questions. Based on the latter, three half-day parallel sessions were held during which the Consortium members systematically gathered experiences and views. The results were documented on flip charts and/or pin walls and later collated and interpreted by the session facilitators.

During the second GNU Extended Consortium conference (Rome, Italy, 4 February 2010), these preliminary outcomes were presented to the members of the Extended Consortium, consisting of stakeholder representatives who operate on the national or sub-national (that is subsidiary) levels and who represent user organisations which do not participate in the GNU Consortium and researchers, policy makers and service providers. After this plenary presentation, dedicated splinter groups were formed where the work was disseminated in more detail and feedback was gathered to add the perspectives of the attending stakeholder representatives¹.

The present document reports the combined outcomes of these efforts. It describes a comprehensive users' perspective regarding the optimal setup of data portals for GMES and related data products or services. It is intended to provide recommendations and guidance to organisations who design and install such portals or modify their existing ones. A sister task follows the present one. Using the identified criteria, a questionnaire was drafted and distributed among the Consortium members, each of whom was responsible for in depth evaluating one or more particular data portals using this questionnaire. A well balanced mixture of portals was evaluated and no evaluation was duplicated. These results will be collated and interpreted and presented in a separate report.

¹ Conference results are available via the project website.

The GNU partner's acquaintance with data systems

The overall acquaintance of the GNU partners with existing GMES and related data portals was investigated to outline the framework of the work presented herein and to document to which portals the below described opinions are related.

The Consortium and Extended Consortium members are primarily familiar with national and sub-national geoportals (e.g. AT, BE, DE, NL, FR, IT). These systems are the ones most frequently used by all GNU participants. Those who collaborate with the EEA (European Environment Agency) frequently use EEA data services, too, such as Reportnet and the Common Data Repository (CDR) for uploading information to the EEA. Further portals well known are GMES geoportals by ESA (European Space Agency) and the European Commission, and portals by EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites). Data are also retrieved from several NASA (United States National Aeronautics and Space Administration) portals. GNU partners frequently use several commercial data portals by data providers focussing on EO (Earth Observation) & GIS (Geographical Information Systems) and, not least, Google Earth.

GNU members primarily obtain data from portals in that they act as clients. However, several of them also act as information providers in that they upload data to national systems. Those reporting to the EEA frequently upload data to the EEA's system, and the Extended Consortium also contains some commercial service providers.

An overview of web-based data systems that certain GNU Consortium or Extended Consortium members are well acquainted with is given in the below list. It reflects the portals and catalogues GNU members were already familiar with owing to their previous work. The list does not reflect any selection made for the present task, nor does it intend to be complete.

- Airbase
- ARPA
- CDIAC
- CENIA geoportal
- Climate.scert.ru
- CORINE database
- Czech geology Survey
- DATA-FED
- DODS
- EEA data service (Reportnet, CDR)
- Eionet
- Enviroportal.sk
- EOLISA
- ESA Rolling Archives
- ESPON
- EU Forest Data Centre
- EU Soil Data Centre
- Eumetsat
- EuroForest Portal
- Forest Carbon Portal
- GEO Portal*
- Geoland.at
- GEOMIS Bund
- GEONODE
- Geoportail IGN-FR
- GISU
- Google Earth
- GSE FM
- INGV
- INSPIRE Geoportal
- JRA
- Land Use Data Centre
- Land.eu
- Meteo.lu
- MUDAB
- NASA GCMD
- NASA Giovanni
- NASA LAADS
- NASA-OceanColor
- NERC data centres
- NLCSK
- NOAA Climatic data centre portal
- OneGeology
- PCN
- Promote webpage
- PRTR
- SACCESS
- Sensorweb
- Sinanet
- SIRIUS Catalogue (Spot Image)
- SISCAL
- Soil Portal
- Umwelt Daten Katalog
- USGS
- WISE

* From June 2008 to June 2009, GEO (Group on Earth Observations) conducted a testing of its portals. Several GNU members participated in this effort.

Strengths and weaknesses of different data portals

The sheer number of existing portals is confusing, and an effort to consolidate and reduce this number would make it easier for users to be well orientated within the supply of GMES. GNU participants show clear preferences for certain data portals over others – with a general tendency of American portals considered more user-friendly than European ones, but with exceptions.

User friendly setup

- *Easily comprehensible user-interfaces* (e.g. NASA LAADS) reduce the time needed to get acquainted with the portal is short, whereas the overly complex structure of certain systems can form an obstacle
- *Easily comprehensible and logic structures and hierarchies* enable users to easily find their way, whereas some portals show a deep layering and many steps are needed to access data (e.g. PROMOTE), or data sets must be chosen from lengthy lists
- *Clear target groups* ascertain that content is customised and users can estimate which data they will encounter, whereas some systems apparently without a clear target group contain too many or too diverse data which makes it difficult for users to keep an overview
- *Customisable and specific query and filter options* help to locate data sets because each user has a different way to search for them; also sound meta-data (see below) are an essential prerequisite to find data sets for specific purposes
- *Drop-down menus* facilitate entries so that the user does not need to exactly know which terms to enter
- *No registration required* (e.g. NASA) or, if needed, a *simple procedure* much speeds up the process to access data, whereas for some systems (e.g. ESA CAT-1), registration is so onerous that it forms a serious obstacle to use them²
- *Compatibility with standard browsers* enables users to access data without encountering software problems, whereas some systems can cause browser security conflicts (e.g. ESA)

² For some of the matters mentioned, registration is needed, especially user-tracking for customising the portal and also for notification of updates, so a sensible approach should be chosen.

Efficiency

- *Batch downloading* (e.g. NASA systems) saves time as opposed to downloading one set after the other
- *Customisation on the fly* (e.g. SACCESS) speeds up repetitive searches
- *Notification for updates* avoids that users visit the portal frequently to look for new data in vain
- *Filling in standardised meta-data via a thesaurus* ascertains correct terminology and spelling and avoids misguided searches (may go along with drop-down menus), because one and the same issue may be termed differently by different operators, which could be especially helpful for non-native English speakers
- *One entry point* (e.g. Enviroportal.sk)³ enables users to visit one portal for different purposes rather than searching several places which is more time consuming
- *Quick looks* (e.g. SPOT Image, NASA systems) much speed up any search because the user can quickly assess whether a particular data set is what they are looking for before actually downloading it
- *New/latest content indication* enables that this can be distinguished from older entries which the user may already have downloaded earlier
- *Most frequently accessed data automatically go up* especially if data are presented in long lists, so that the most frequently searched for data are quickly accessed (other ways of sorting, e.g., alphabetical should of course also be possible).
- *User tracking* can be very helpful for customising the system and adapt contents to what is most looked for (but inevitably requires registration⁴), and, not least, to identify gaps in current data delivery (failed searches)
- *Stable system* which is mostly the case, but some systems frequently frustrate users by going down during searches or downloads
- *Fast browsing and high download speed* considering typical internet connections, whereas some systems tend to a slow response time, even for browsing (GEO Portal, GEOMIS Bund)
- *Resuming large downloads in case of transmission breaks* can save time in that after a break the download continues rather than re-starting from the beginning
- *Checksums available to the user* verify the downloaded file is intact and can also show that the data are not modified.

³ This will be mutually exclusive with the idea of portals focussing on to a clear target group.

⁴ Unless the portal provider wants to rely only on the IP addresses, which may tell the affiliation of the user but no more details.

Economy

- *Free and unlimited data access* (several examples) improves the attractiveness of any system, especially for users working in the public domain with limited budgets, whereas systems offering primarily or only restricted data are less appealing
- *Cost info given at top-level* because obviously not all data can be made available for free, but cost-transparency is in any case essential, and being confronted to costs only at a later stage of data searching will frustrate any user

Availability and standard of the data

- *Direct access to located data*, that is not only learning that these data exist, but either obtaining them directly from the system or being linked to a direct point of access forms a key feature to enable users to integrate them in their own applications; another option would be inline linking⁵.
- *High level standardised products* as described in the deliverable *GNU GMES Product Quality Criteria*, whereby one should also consider that overly ambitious specifications can form a barrier to enter/register products
- *Quality checks at upload of content*. ensure that users know they can trust the contents of the system and can limit their own quality checks to a minimum
- *Rapid access to marked unrestrained (unchecked) data* (e.g. SACCESS) is useful in certain cases where speed of delivery of very recent data is more important than elaborate quality checks
- *Large amounts of data for download* and ideally complete data sets and not just test sites ascertain that the particular information the user looks for is present in the system
- *Timeliness/up-to-date state of contents* (e.g. NASA LAADS) ensures that information reflects the current (environmental) status which the user works on, whereas some systems show a significant time lag or even out-of-date contents (e.g. MUDAB) which can significantly reduce the data sets' utility
- *Adequate meta-data* specify the framework in terms of time, area, resolution, quality assessment, production methods etc. (cf. *Essential meta-data* below); unfortunately certain systems offer data without or with incomplete meta-data (e.g. UDK) so users can hardly judge whether the data fit to the intended purpose

⁵ An object such as an image which physically is stored in one site is used by a web page of another site, the user does not directly notice where the object is located.

Help systems

- *Clear user guidelines* make certain that users get the most out of a particular system
- *Detailed upload instructions* online or even specific education for individuals who upload data will avoid erroneous entries and nuisance
- *Embedded user support*, for instance written instructions, tutorials, examples of typical searches, FAQs (Frequently Asked Questions), possibly even telephone support (e.g. Rapid Eye, UDK) help users to learn to operate the system
- *Translation for non-English websites* ensure that data from the country of origin can be accessed by all, currently some mono-language websites hamper this access

Advantages of obtaining data-products through portals

From the users' perspective, there are several reasons why obtaining data-products or services via portals is generally considered advantageous in comparison to obtaining them directly from one producer. Several of these points go along with certain strengths of portals listed above.

Knowledge

- *Directory function*, that is the portal lists in an easily searchable way which data sets/services currently exist and are accessible for a particular purpose
- *Becoming aware of data sets* not yet known to the respective user and possible serendipity of data discovery
- *Comparison and selection* of different products from different sources rather than finding only products from one producer

Communication

- *Opportunities to use portals as a communication tool* between users and producers
- *Opportunities to get in touch with peer users* if public feedback mechanisms exist

Ease

- *One stop shop* or even first-stop-shop, that is saving time by visiting only one place rather than collecting data from several sites
- *Combination of many layers/scales etc.*, simply owing to a range of products available larger than one producer could present
- *Less to learn for users*, because they have to get acquainted with one setup rather than that of a range of individual providers' web sites

Quality

- *Controlled QA* (Quality Assessment) by a recognised authority, so data can better be trusted
- *QA harmonisation*, that is the same quality standards can be applied to different products
- *Hallmarking*, giving products or services a stamp of authenticity by a recognised authority
- *Consistent meta data* between different products, achieved e.g., through an entry form to guide individual metadata authors (or through centralised metadata editing)

Incentives and barriers to register products in a data system

Although for users, data portals usually are advantageous, data providers⁶ face a rather differentiated situation involving both incentives but also barriers which will influence their decision whether they should use a portal or not. Especially possible barriers should be considered when organisations design data systems and expect that data providers will want to register their data. Several of the below incentives are interrelated or paired with particular barriers mentioned, and the resulting effects (e.g. increased or decreased visibility of the respective data providers) will result from the particular portal setup.

Incentives

- *Business opportunity*, because making data available via portals may lead to additional sales
- *Publicity* by being present in portals which generally increases the visibility of the provider and their data and services, thus provide an advertisement and may cause users to contact this provider directly for further services
- *Permanent (24/7) availability of the data* independently of office hours, vacations, national holidays, and the like
- *Adding to users' peer pressure*, in that users notice that other users take part and thereby may feel a me-too-effect (likewise, the fact that a provider participates in a data portal may result from providers' peer pressure in analogy)
- *Cost saving* in data provision may potentially arise from distributing data via portals rather than distributing them through own channels or personally
- *Support for dissemination* of ongoing or past project work which has led to data being available for distribution
- *Feedback by users* may help providers to permanently improve the offered data sets in case feedback loops are enabled by the system
- *All of the above described advantages of data portals as seen by users* should be considered incentives for data providers to use portals, because these advantages will eventually lead to an increased use of their data and services

⁶ GNU is a network of GMES data users, but as mentioned above, several GNU partners also enter data in national data systems. Moreover, the two SMEs and the two EO research organisations in the network also act as data providers, and some commercial service providers are represented in the Extended Consortium.

Note: A general incentive could simply be to encourage or even require data produced using certain funding mechanisms (e.g. FP7) to be registered in particular portals.

Barriers

- *Abundance of data portals*, because the sheer amount of them makes it difficult to focus and to ensure that the data will be found at all
- *New obligations*, such as the need to permanently keep records up-to-date and to ensure continuity of provision
- *Additional work without adequate return*, in that publicising services may just lead to an increase in enquiries rather than in revenue
- *Increased efforts* in comparison to direct distribution through the need to register and through adapting data sets to required standards such as metadata and formats
- *Unknown future system-demands* in that the portal may require the input of additional (meta-)information at a later stage e.g. when it is revised by its operators
- *Uncertainty of costs* if long-term charges are not clearly described
- *Reluctance to publicly provide details about production procedures* which may be needed to enable users to understand the data quality but can at the same time be used by competitors
- *Fear for newly arising competition* in that data sets are directly compared with those of competitors
- *Risk of losing control of portal contents* and possibly having Intellectual Property Rights undermined in this way
- *Direct access to the provider's data* through e.g. Web Feature Services (WFS) and/or Web Map Services (WMS), which may cause conflicts regarding security related data
- *Fear to be exposed to unknown Quality Assessment procedures* including a risk to be unjustly criticised or even have one's data and services mis-interpreted
- *Possible weaknesses of the portal may be attributed to the provider* if the portal is difficult for users to work with, or the system does not sufficiently consider existing user requirements
- *Reduced visibility* of provider as the service information is pooled with that of many others

Essential meta-data

Meta-data are indispensable to find suitable data products or services and to understand whether they will fulfil the respective users' expectations. The below list of potential meta data is intended as a general guidance. It is not expected that all these meta-data will be gathered for each data set, which would mean an unreasonable effort for the respective data provider, but a sensible choice should be made, because, as mentioned above, overly ambitious metadata requirements likely form a barrier for registering products. A relatively small number of meta-data, but with very high quality and in an absolutely consistent format will be more helpful for portal users than a large amount of low-quality and inconsistent meta data. The meta data should be compliant with particular standards (such as ISO 19115, ISO 19139, Dublin-core⁷) to facilitate interoperability between different catalogues and to make it easier for users to work with several catalogues. Those metadata schemes should be chosen which are suitable for (established within) the field of the respective data or services. In any case, meta-data should clearly be defined in accord with requirements defined through INSPIRE⁸.

General product info

- *Topic/sector/category*
- *Product and/or service title*
- *Abstract*
- *Keywords* including thesaurus or drop-down to ascertain coherence, better than open keyword entries
- *Policy driver*, for instance an environmental directive or indicator to which the data are related
- *Intended purpose of the data set*.⁹
- *State*, e.g. research, pre-operational, or implemented

⁷ Dublin Core Metadata Initiative“ (DCMI), 1994

⁸ Infrastructure for Spatial Information in the European Community

⁹ An important concept of the deliverable *GNU GMES Product Quality Criteria* is the *Fitness for purpose*: The quality of a data set or service can be judged only within the framework of its intended usage.

Quality info

- *Validation standards* applied and access to their documentation
- *QA/QC* (Quality Assessment/Quality Control) procedures
- *Service Level Agreement*, if any was signed by a user organisation for this product
- *User rating*, optionally an integrated mechanism enabling users to publicly indicate their reception of the particular data set (see below, *Integrated user feedback*)
- *Time stamp* automatically updated as a measure to show whether the page is used or outdated

Resolution and Accuracy

- *Resolution* (scale): spatial/spectral/temporal
- *Accuracy*: thematic, spatial
- *Reference year(s)*

Availability and applicability

- *Availability*, geographical and temporal
- *Completeness of data*, in terms of spatial or temporal gaps and of accessibility of the complete data set through this portal (if only parts are copied to the portal, the user would have to download the actual data set from another place)
- *Coverage*, geographical and temporal, or whether only test sites exist
- *Geographical applicability* of the method (e.g. by biogeographical region or another geographical reference)

Inputs, Methods and Outputs description

- *Technology and tools* (hard- and software) required to read the data
- *Physical/chemical parameters/attributes* including definition and description
- *Methods*, including description of data manipulation procedures and indication if these methods are published/unpublished/proprietary
- *Data types* for input, intermediate work, and output

Product previews

- *Quick looks*
- *Real examples*
- *Links to outputs* for which the services were used (e.g. reports)

Service provider and IPR info

- *Service provider(s)* and link to their webpage (avoids the need to have a description of the service provider in a separate metadata form)
- *Contact person*/expert at service provider
- *Copyright information* on data and methods
- *Data owner* or holder

User info

- *User information* and link(s) to their webpage(s) may be helpful to get into contact with other users who apply the product and to exchange experiences

Costs

- *Costs* and (long-term) pricing information
- *Changes of cost categories* if dependent on the kind of usage

Integrated user feedback

For users who consider downloading or even purchasing data, it is in any case very helpful to be informed about experiences other users may have made with these data. Thus, a functionality that enables publicly entering user feedback, evaluation or ranking of data products in a catalogue could be integrated in any data portal. This approach has already become an established standard for commercial online shops such as *Amazon*, *Booking.com*, or the *iTunes App Store*, in which users can both rank products using a star-system and also provide written comments. These user evaluations can be accessed individually and are also provided as a statistical (number and average) value.

A user feedback mechanism integrated in a data portal could become a crucial component of influencing in a truly user-driven way which contents it will present. However, possible legal consequences must be considered, which would occur when users give incorrect or biased statements about products or services. Measures to avoid such cases would involve registration and public naming of the individual users who provide product evaluations, and an explanation how and under which cir-

cumstances the product was being used. Moreover, user feedback could sensibly be guided by a standardised feedback form (e.g. a simplified version of a service utility assessment as applied within GSEs (GMES Service Elements)). A fair and transparent evaluation procedure is in any case essential, because otherwise data providers could not be expected to register/upload their data sets in the respective portal. In the above mentioned online shops, products will typically be ranked by large numbers of users, so that individual misguided (positive or negative) statements will statistically not make much of a difference. This situation is completely different within GMES where each individual ranking would have to be suitable to be taken serious. Moreover, some online shops have a very strong (or even dominant) market position, and sellers do not have much of a choice whether or not to use them. This situation is again different for GMES where a provider could simply remove a negatively evaluated product from one portal and place it into another one – or simply not use portals at all.

Portal authenticity

The opportunities provided through current web technology would enable basically anyone to set up a data portal and quickly fill it with contents, e.g. by combining and modelling contents of other sources. Therefore, users appreciate an authentication of the portal contents, including a clearly indicated origin of the data or services (a real organisation), explicit explanation of the production chain, information about an independent validation process, or approval by a public authority (as data owner, manager or holder). A future possibility might be the often discussed *GMES Quality Seal* which could be provided by the European Commission and ESA and declare that a portal or a particular data set provides trustworthy contents. This would avoid that users must in detail test each data set themselves which they consider to use in the future. Such an established quality seal could also be helpful for young SMEs (Small or Medium Enterprises) which provide valuable products, but are not yet well known, so users are still hesitant to trust these companies.

A weakness of a quality seal would be that it could ensure only data authenticity, data integrity and source authentication of data but not of subsequent changing of data. The INSPIRE principles could reduce such problems resulting from the changing of data and services through the rule that *data should be collected once and maintained at the level where this can be done most effectively*.

Beyond portals and catalogues

Considering the already vast and quickly growing and changing amount of data, portals which are set up following current data management concepts may very soon not anymore be the most suitable way to provide access to GMES data and services. Already today, the concept of establishing large data repositories is increasingly superseded by the approach to keep data close to their origin and make them accessible anywhere through an effective data network (e.g. SEIS, Shared Environmental Information System).

Portals could even, in a Web 2.0 fashion, be replaced by a market-place like structure without any central data entry and with very basic rules. Data providers could enter or link to any data products as they see fit, comparable to *GMES Yellow Pages*. In a further developed system, users would respond to posted products with evaluations and ranking. Such a system would become mostly self-regulatory and become somewhat comparable to the current *eBay*. An even more open approach could be a *GMES search engine* linked to INSPIRE metadata and catalogue services standards. The GMES data products could then be stored anywhere within the World Wide Web, but they would be tagged with a centrally defined set of keywords so that they can be located.

This report and all other reports by GNU including those mentioned in the present document can be downloaded from the project website www.gmes-network-of-users.eu