



Southern European Seas: Assessing and Modelling Ecosystem changes



Project no: 036949

Project acronym: SESAME

Project title: **Southern European Seas: Assessing and Modelling Ecosystem changes**

Instrument: Integrated Project

Thematic Priority: 6.3 Global Change and Ecosystems

SESAME Data Policy

Start date of project: November 1, 2006

Duration: 48 months

Project co-funded by the European Commission within the Sixth Framework Programme (2006-2010)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	X

SESAME DATA POLICY

In accordance with the FP6 Rules governing the Intellectual Property Rights (IPR) of the funded projects clearly stated in SESAME documents [SESAME Contract (Articles II.9 - Confidentiality; II.10 - Communication of data for evaluation, impact assessment, standardisation purposes and communication of information beyond the research community; and Part C - Intellectual Property Rights (II.32 - Ownership of knowledge, II.33 - Protection of knowledge, II.34 - Use and dissemination, II.35 - Access rights)) and SESAME Consortium Agreement (Article 8- Intellectual Property Rights (8.1 General Provisions; 8.2 Confidentiality; 8.3 Publication; 8.4 General principles regarding Pre-Existing Know-How (PEKH); and 8.5. General principles applicable to Knowledge)] SESAME is thereby identifying the rules of its data policy.

The aim of this policy is to facilitate operation and use of the SESAME data base by the research community of SESAME. This policy recognises the benefits of providing free and open access to good quality data from environmental and socioeconomic sciences in order to achieve SESAME's goals as stipulated in the DoW and the Consortium Agreement as well as for future use in global change studies, research projects, and operational services such as portals and search engines. The steering committee of SESAME encourages the widest possible use of the data base, in order to best realise the value of the SESAME data collection during the project as well as after its completion.

Principles

- The SESAME data base is the central archive to all SESAME data. The system supports free and open access to its content by the SESAME's partners. In the future similar access will be extended to research and education communities in non commercial activities in line with the EU rules.
- For any data, provided by SESAME, the format, content and documentation of a data set must ensure its most widespread and easiest use by the scientific community.
- Users of data downloaded from SESAME data base are obliged to properly use the data set citation and/or quote the related reference.

Data provision

In agreement with all partners and following the work program and deliverables as outlined in the SESAME contract, any scientific primary data (1) already existing and being of

relevance to SESAME or (2) produced through research within SESAME should be archived in the SESAME data base. This includes data resulting from expeditions, on board or post-cruise measurements, documentations and publications, formatted in machine readable form.

- Data archiving includes:
 1. Metadata(*) of expeditions, stations and samples;
 2. Scientific primary data from a) archives, b) expeditions, c) publications;
 3. Metadata related to the primary data (2);
 4. Products resulting from compilations and interpretations of primary data (e.g. model outputs).
- Chief scientists are requested to send cruise reports including a station list to the project management office (Dr Eleni Kaberi, email: ekaberi@ath.hcmr.gr) within 45 days of the end of the cruise, and submit metadata for each cruise's dataset (for example CTD dataset, bottle dataset, plankton net dataset, etc.) to the SESAME data portal. Station labels as published in the cruise report's station list must remain the same at any time when used in data submissions or publications.
- The data manager maintains a dictionary of parameter definitions with unit, to be used as the agreed standard for all project data. Parameters are grouped into categories according to their related scientific field. Data submissions are required to use parameters and units as defined in the dictionary. New parameters are defined by the data manager on request.
- Data are archived in a relational database, georeferenced in space and time; if a data set is very large or, for certain format reasons, must have a proprietary format, it is archived as a binary object in a file system with a metadescription in the data base, linked to the file.
- As soon as new data become available and are validated, the providers are requested to submit the data in agreement with the import format. Any type of data should always be accompanied by a description (metadata) allowing future users to understand and process the data at any time.
- Existing data will be integrated to the SESAME data archive if provided on request to former projects by the data management.
- The granularity and import format of data sets is defined in agreement between the principle investigator (PI) and the data manager.
- The export format in principle is tab-delimited ASCII, headed by metadata fields according to ISO19115, GCMD-DIF and DublinCore standards.

Quality assurance

- Data submitted for archiving should be documented properly; documentation is archived together with each dataset.
- The scientific quality is in the responsibility of the PI/author. The required fields for data documentation (quality flags, precision of values, and documentation of methods) are available in the data format descriptions which are accepted by SESAME partners.
- Technical quality control, i.e. completeness of metadata, consistence of formats and correctness of download is in the responsibility of the data managers.
- After import, the PI/author is requested to proof read data sets on the Internet and submit corrections to the data manager.

Access and Publication

- According to EU data policy, all data collected during the lifetime of the project are made public two years after the termination of the project (Article II.34 (Use and Dissemination) of the Contract)¹; regulations for certain data collections may differ in agreement between coordinator, partners and funding organization. Confidentiality issues are regulated by the SESAME Contract and Consortium Agreement (Articles II.9 and 8.2, respectively)².
- The use of any scientific data for publication should follow the EU rules as stated in SESAME's Publication Strategy (http://sesame-ip.eu/scientist/partner_publi.php). Data providers should be referenced and/or acknowledged. Datasets would be flagged in the data base with 'acknowledgement-reference details' by the providers.
- When submitting data to the SESAME data base, data providers should define the "availability" of datasets according to the following categories as stated in the data base:

Originator: The dataset is password protected and only available through contact to data originator.

Partner: The dataset is available to all SESAME participants.

¹ Article II.34 - Use and dissemination of the Contract states:

'If dissemination of knowledge would not adversely affect the protection or the use of IPR and confidentiality issues, the contractors shall ensure that knowledge is disseminated within a period of two years after the end of the project. Should the contractors fail to do so, the Commission may disseminate the knowledge. In so doing, the Commission and the contractors shall take particular account of the following factors: a) the need to safeguard intellectual property rights; b) the benefits of swift dissemination, for example in order to avoid duplication of research efforts and to create synergies between projects; c) confidentiality; d) the legitimate interests of the contractors'.

² For more details on confidentiality issues also refer to Articles II.10 and Part C (Intellectual Property Rights) of the EC Contract and Article 8 (Intellectual Property Rights) of the SESAME Consortium Agreement.

Public: The dataset is available for any authenticated individual with sufficient credentials to satisfy SESAME access logging requirements.

- Data may also be submitted if validated but not yet published. Unpublished data are password protected by default. Password protection for published data is set on request for a moratorium period. The originators may decide to withdraw unpublished data from the archive. Metadata are always freely accessible.
- In case that the submitted datasets are characterised as “public” they would be also in PANGAEA information system.
- Partner institutes and data providers agree that data archived in PANGAEA are made publicly available on the Internet (e.g. portals, search engines, library catalogs, GIS) without further notification. Persistent identification, data publication and widespread distribution are performed by the networking functionality and web services of PANGAEA.

Operation

- Long-term availability of data is ensured by IOLR, responsible for the technical operation and the consistency of the content through its connection to SESAME web page.
- The Backup of the data inventory is in the responsibility of the computer center of IOLR with daily incremental backup and weekly full backup in two mirrored drives, located in different buildings.
- Data flow is organized from the work packages via the data manager Isaac Gertman to the archiving facility at IOLR, monitored by the project management and supervised by the data manager.

(*) Depending on the level of processing scientific factual data can be differentiated between raw data, primary data and secondary data (higher level products). Raw data are provided by a measuring system and are not processed; scientific primary data are resulting from the processing of raw data and are the basis for scientific interpretations and publications. Primary data have the highest priority for archiving; the related raw data files are added if appropriate. Secondary data are higher level products resulting from compilations and interpretations of primary data, i.e. maps, profiles, statistics, images, films, graphics or any material produced for education and outreach. All information describing any of these three data types are metadata.