EGI Scientific Discipline Classification Virtual Team

Final Report

Authors/Members:

Sy Holsinger (EGI.eu), Sergio Andreozzi (EGI.eu), Gonçalo Borges (LIP), Marios Chatziangelou (IASA), Claire Devereux (STFC), Iván Díaz (CESGA), Maciej Filocha (ICM), Cyril L'Orphelin (CNRS), Geneviève Romier (CNRS), Alvaro Simon (CESGA), Jelena Tamulienė (VU)

Abstract:

This report provides a proposal for a new classification of scientific disciplines for EGI, methodology used, technical issues and a set of recommendations to be endorsed by EGI management for implementation.

Table of Contents:

Recommendations **2**
1 The Virtual Team **4**
1.1 Overview *4*
1.2 Members *4*
2 Methodology and Research **5**
2.1 EGI Classifications and Uses *5*
2.2 Available Public Classifications *6*
3 EGI Tools: Technical Analysis and Considerations **6**
3.1 Accounting Portal *7*
3.2 Applications Database (AppDB) *7*
3.3 Customer Relationship Management (CRM) *8*
3.4 Operations Portal *9*
3.5 Training Marketplace *10*
4 Community Feedback and Public Comments **11**
4.1 VO Manager Survey Feedback *11*
4.2 Public Comments *12*
5 Conclusions **12**
6 Proposed New Classification **13**
6.1 Overview *13*
6.2 Detailed Classification *14*
7 References **20**

Recommendations

The SDC VT performed an in-depth analysis of public classifications, current uses within EGI tools (Accounting Portal, Applications Database, Operations Portal, Training Marketplace, Customer Relationship Management) and have defined how a new classification scheme could be implemented and identified related issues. It is out of the scope of the VT to have implemented such changes, but to provide a set of recommendations for both EGI management decisions and tool operator implementation moving forward. The 5 recommendations are as follows with further details provided in the rest of the report:

Recommendation 1

- EGI management to prioritise work based on the estimated effort required to make the technical changes and mandate the tool operators to implement them.

It was out of the scope of the VT to make effort and resource decisions for where and how much to dedicate to the technical changes. Therefore, it is a recommendation from the VT for the EGI management to evaluate closely the required changes, availability of effort and prioritise the necessary resources to implement the classification (See 3). Effort for implementation is available in PY4 for the Accounting Portal, AppDB and Training Marketplace, while the CRM and Operations Portal would need to prioritise other work accordingly. One option to support the work and handle any potential issues that may arise could be to create a new Virtual Team dedicated to implementation of the classification.

Recommendation 2

- EGI management to identify the most appropriate location for central hosting for automated processing of a master classification

In order to distribute the common scientific classification tree among the EGI services and keep them up-to-date on potential new changes (new additions, updates or even deletions), a single end-point needs to be created that will offer the EGI scientific classification to the rest of the EGI services over a simple, developer-friendly, stable/reliable and well defined API.

An indicative sample response, in a pseudo-XML format, could be:

- <discipline>
- <discipline id=”7” level=”1” parentId=”8” state=”**active**” updated=”XXX”>Humanities</discipline>
- <discipline id=”7.1” level=”2” parentId=”7” state=”active” updated=”XXX”>History and Archaeology</discipline>
- <discipline id=”7.2” level=”2” parentId=”7” state=”active” updated=”XXX”>Languages and literature</discipline>
- <discipline id=”7.3” level=”2” parentId=”7” state=”**obsoleted**” updated=”XXX”>Arts</discipline>
- <discipline id=”7.1.1” level=”3” parentId=”7.1” state=”active” updated=”XXX”>Archaeology</discipline>
- …

Since the EGI AppDB (appdb.egi.eu) service makes a significant use of technologies related to APIs, actually the entire service has been built on the top of a RESTful (XML and JSON) API, the service should be considered as one of the potential candidates of holding and offering such an API to the rest of the services. A user interface accessible only from members authorised to perform changes on the scientific classification schema could also be implemented.

Recommendation 3

- Tool developers should implement to level 3

Using a connected structure, selecting level 3 provides scalable information based on the communication channel used (e.g. presentations, statistic reporting). Therefore, encouraging users to select a scientific discipline from level 3 would provide the most useful information. It is understood that the level of detail may be difficult for some tools to implement and visual representation may be difficult. If some tools provide only level 2, there is a chance that some information (even information already currently available) may be lost. It is understood that the Training Marketplace will only integrate up to Level 2 as Level 3 is not required information for their purpose, however this would not be the case of tools such as the Operations Portal.

Recommendation 4

- Tool developers should enable multiple selection of disciplines

The added functionality of "multiple selection" is one of the key recommendations made by the VT. Multiple-selection provides a full understanding of what disciplines are being covered by a single VO, which also avoids a catch-all classification of "multidisciplinary". Each tool is to ensure that this can be implemented. Explicit instructions should be included to help users in each tool (e.g. catch-all VOs to select all Level 1 categories). This will ensure potential users find these VOs when searching for their discipline.

Recommendation 5

- EGI management to identify the responsibility of change management and process ownership

One of the main goals of the VT was to ensure that the agreed classification would not need to be frequently changed or would become obsolete within the short-term leading to another massive overhaul. However, as the community evolves, the classification will still need to be periodically reviewed. This requires a defined process and ownership to ensure that the classification is not only maintained, but reviewed and updated following an agreed systematic process. This could be done through an annual review process with a reoccurring Virtual Team with EGI.eu as owner.

1 The Virtual Team

1.1 Overview

EGI is a multidisciplinary e-Infrastructure where users belong to a variety of different scientific disciplines. EGI needs to categorise these users by disciplines through a number tools (e.g. Accounting Portal, Applications Database, Operations Portal, Training Marketplace, Customer Relationship Management) as well as communicate externally as to who is using the infrastructure (e.g. funding agencies, current and potential new user communities). Although a legacy classification was inherited from the predecessor project EGEE (Enabling Grids for E-sciencE), different tools have adopted different classifications and the expanded user base has many Virtual Organisations (VOs) falling into the "other", “multidisciplinary” and “Infrastructure” categories (approx. 50%) with the rest spread across only seven disciplines. As EGI has continued to expand the usage of the infrastructure through an open ICT ecosystem, the current classification is no longer indicative of the current usage. In addition, it does not allow for the accounting of new communities, therefore, it has become essential to agree on a common, coherent classification that is not only consistent across all tools, but allows for smooth inclusion of both current and future user communities.

The Scientific Discipline Classification Virtual Team [R1] (SDC VT) was formed to provide a proposal for a new classification of scientific disciplines for EGI that is verified with the VO managers, EGI tools operators and NGI International Liaisons (NILs). This activity was initiated to implement a recommendation from the Scientific Publications Repository VT adopted by the Council [R2]. The main activities of the SDC VT were to:

- Identify all possible uses of disciplines across EGI.
- Research publically available classifications.
- Define an aggregation of scientific disciplines.
- Understand the technical implications of integrating a new classification scheme.
- Present the proposed list for comments and recommendation by VO Managers, EGI tool operators and NILs.
- Submit an agreed and verified classification to EGI Management and Council for approval with a set of recommendations moving forward.

The following sections provide further details regarding each of these activities as a final report of the Virtual Team.

1.2 Members

The Virtual Team and its objectives attracted a wide range of participants with varying expertise. A representative from each EGI tool was present as well as several NGI NILs and representatives. EGI.eu staff provided the overall coordination and management of the Virtual Team. The members are as follows:

- Sy Holsinger, EGI.eu / VT Leader
- Sergio Andreozzi, EGI.eu
- Gonçalo Borges, CRM
- Marios Chatziangelou, AppDB
- Claire Devereux, UK NIL / TMP
- Iván Díaz, Accounting Portal
- Maciej Filocha, PL NGI
- Cyril L'Orphelin, Operations Portal
- Geneviève Romier, FR NIL
- Alvaro Simon, Accounting Portal
- Jelena Tamulienė, LT NIL

2 Methodology and Research

In preparation for the Virtual Team, background research was conducted in order to better articulate the issues and formulate the rationale for the Virtual Team. An investigation was made into the various EGI tools, how and what scientific disciplines were being presented, coupled with what public classifications where already available. The research outlined how diverse scientific disciplines were being used as well as a number of public classifications that have been attempted by various organisations. It was important for the Virtual Team to not carry out work that had already been done and focus on applying the most appropriate classification to EGI balancing the level of detail with ease of use. The Virtual Team worked collaboratively through a Google Spreadsheet for tracking the work, providing comments and feedback and for ultimately defining the final classification for review [R3]. Further details are provided in the following sections.

2.1 EGI Classifications and Uses

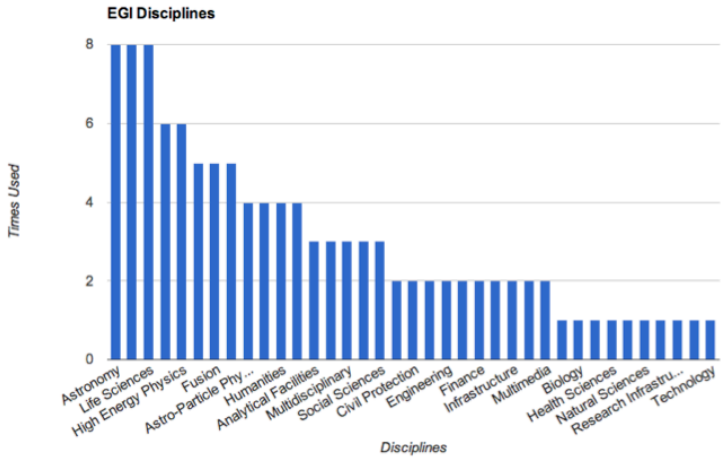
The first step of the Virtual Team was to understand what EGI tools use or present scientific disciplines and how were they being classified. There are a total five tools (Accounting Portal, Applications Database, Operations Portal, Training Marketplace, Customer Relationship Management) and a number of communication channels such as EGI presentations, reports, use cases and Virtual Research Communities that have specific scientific disciplines as well. These channels mentioned 38 total different disciplines with varying use. Only three channels use the same: Operations Portal, AppDB and the EGI Compendium. However, the AppDB does include more than 300 sub-disciplines, but with no connected structure.

The major issues have been that the majority of tools are very specific when defining the disciplines with limited or no flexibility, therefore, many fall into the "other" category. The "multidisciplinary" and "infrastructure" categories are also commonly used. The reasons differ, but are mainly because a single VO can cover a number of different disciplines, and some VOs have been set up to facilitate training and demonstrations, application porting, or be a "catch-all".

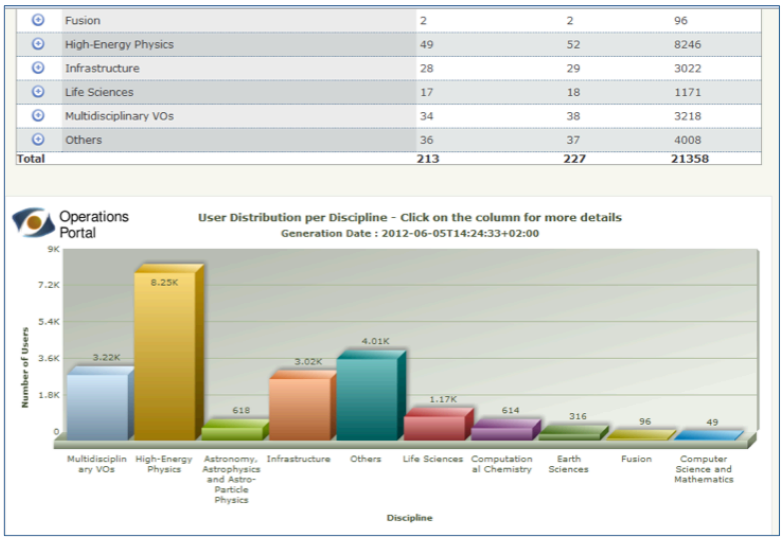
This makes it difficult to:

- Understand from where the actual users come.
- Provide consistent communication and accurate statistics.
- Easily account and integrate new user communities.

The following figure graphically represents the number of times different scientific disciplines are used in EGI tools and communication channels.



The below image is currently how usage statistics by scientific disciplines is being presented by EGI (generated by the Operations Portal). It graphically represents how the majority of users are being grouped in generic, unspecified categories.



	Fusion	2	2	96
	High-Energy Physics	49	52	8246
	Infrastructure	28	29	3022
	Life Sciences	17	18	1171
	Multidisciplinary VOs	34	38	3218
	Others	36	37	4008
Total		213	227	21358

2.2 Available Public Classifications

As mentioned, it was important to not try and “reinvent the wheel” when looking to better classify scientific disciplines. Twelve public classifications were ultimately analysed: EuroStat; Cordis; ESFRI; Wikipedia; Google Scholar; SourceForge; Organisation for Economic Co-operation and Development (‘Frascati’); Dewey Decimal; Library of Congress; Universal Decimal Classification; Australian Bureau of Statistics; Dutch Basic Classification.

The proposal offered in Section 6 uses the "Organisation for Economic Co-operation and Development" also known as the “Frascati Fields of Science and Technology (FOS)” classification as a baseline [R4]. In the opinion of the VT, it provided the best classification of scaling level of detail from general field of science to specific sub-functional fields that most appropriately reflected the disciplines currently involved in EGI and allowed for easy integration of future communities as well. In some areas, the classification of the sub-disciplines was vague and therefore the related Wikipedia page for academic disciplines was used [R5], especially in the area of Computer Sciences.

3 EGI Tools: Technical Analysis and Considerations

One of the main issues that held back this work in the past was the difficulty in implementing any classification changes to the tools themselves. The first issue would be how to map the old disciplines being used to the new classification. The second would be the required changes to the functionality of the tools from the user perspective, such as the addition and required feature of “multi-selection” (i.e. allowing users to select more than one discipline as appropriate). Each tool has therefore been analysed separately as not only did each tool use different classifications, but each use different technical solutions for providing the tool. The following sections detail each of the technical analysis and implications by tool as well as estimated required effort for implementing the new classification and availability of effort within the final project year of EGI-InSPIRE.

3.1 Accounting Portal

Currently, the Accounting Portal only has one level of classification (Astrophysics, Computational Chemistry, Computer Science and Mathematics, Earth Sciences, Fusion, High-Energy Physics, Infrastructure, Life Sciences, Multidisciplinary VOs, Others Disciplines, Unknown Discipline).

The mapping between VOs and disciplines is retrieved from the Operations Portal, so there is a strong dependence in that hierarchy. A multi-levelled hierarchy would mandate a change on the XML interface with the Operations Portal to add additional fields.

These classifications are available as a drop-down JavaScript menu in the left pane or as checkboxes on the VO Discipline view. This view already supports the multi-selection of disciplines and the aggregation of non-selected disciplines in the "Other" group. This functionality would be easy to extend to further sublevels.

Visually, some of the names in the left pane are expected to be truncated. As the space in the left pane affects all the views in the portal, perhaps JavaScript code or passive pop-ups to show the complete name on mouse rollover would be needed. For the forms on the VO Discipline view, no problems are expected.

Estimated Implementation Effort:

- 8-10 working days with strong dependencies with the Operations Portal. Would need to do it after their implementation, at least the feed part.

Availability of Effort and Priority Level:

- Since there is an existing maintenance task for the Accounting Portal, the development can be done as part of usual incremental updating and maintenance of interfaces with other tools and services. It would not be top priority, and as stated above, depends on external tool support and availability.

3.2 Applications Database (AppDB)

Based on the current mapping available at the "Mapping (AppDB1)" Spreadsheet, there are two issues:

- ***Issue 1***: There are a couple of "multidiscipline" disciplines that are split into individual scientific fields into the new/proposed schema. For example:

Original (old) discipline values	Proposed to be split into the scientific fields:
Astronomy, Astrophysics, Astro-Particle Physics	Astronomy Astrophysics Astro-Particle Physics
Computer Science and Mathematics	Computer Science Mathematics

As of 21 Feb 2013, there were 60 (13% of the total) software items that are associated with "Astronomy, Astrophysics, Astro-Particle Physics" and 58 (13% of the total) software items that fall under the "Computer Science and Mathematics".

Splitting the (old) disciplines into more than one (new) scientific fields, means that:

A. All the software items that used to be associated with the old discipline, will be mapped with **every** scientific field defined

B. Is it worth requesting a group/VT/body, having as main mission to review the metadata and map these software items one-by-one in a more proper and scientific manner (no automated process included)?

- ***Issue 2***: There are a couple of general/catch-all disciplines in the old schema, i.e. Infrastructure, Multidisciplinary and Others, which are not associated with a specific Functional Field of Level 2.

Simply mapping these along with the VOs could be enough for services like the Operations Portal where the VO is a mandatory value but it is not enough for the AppDB since the VO is an optional field.

Current distribution:

- Infrastructure = 5 (2% of the total)
- Multidisciplinary = 26 (6% of the total)
- **Others 108 (24% of the total)**

The issue is how should the mapping of the software items proceed under the disciplines mentioned above.

As far as the sub-functional field:

- ***Issue 3:*** It will be quite difficult to implement a user interface that will offer all these options to the end-user in a friendly and handy manner. However, for the mapping process, no problems are foreseen, as there is a one-to-one relation between the old sub-disciplines with the new Sub-Functional Fields of Level 3.

Foreseen solution:

An initial mapping of the current categories to the new classification was done in order to better analyse how to move forward. As the developers of the AppDB cannot perform any software item specific adjustments, a couple of scripts will be developed to do the initial migration/mapping and leave the refinement to the owner of the entry. The AppDB policy considers disciplines as a mandatory field, therefore any change that the entry owner would like to do will also have to specify the scientific classification values of the owned item as well, gradually reducing any unspecified entry.

As far as the migration/mapping phase concerns, a two-step process will be followed:

1. Run a script and change the values from the old to the new schema.
2. Inform the users about that change and invite them to log in and make whatever additional refinement they would like.

Estimated Implementation Effort:

- Total implementation = 30 days
    - 20 days up to level 3
    - 10 days developing the users interface
    - 2-3 working days to do the mapping (based on what it is described above)

It is worth to mention that the same developments will also cover the multilevel operational categorisation into the system. Thus, with the given developments/effort, more than one feature is covered. An indicative example of the "operational" categorisation is as follows: Applications, Middleware products (clients, compute, data, information, operations, security, storage), Science gateways (frameworks, instances), Tools, and Workflows.

Availability of Effort and Priority Level:

- There is effort available in the final year of the project to implement the new classification. Priority can be given as directed by management decisions, even immediate if necessary.

3.3 Customer Relationship Management (CRM)

Given the proposed mapping for scientific disciplines in EGI CRM, there are some technical concerns that have to be taken into consideration during the migration/implementation phase:

- ***Issue 1:*** The proposed mapping implies that some of the current scientific disciplines are divided into two or more. Pragmatically, there is no automatic way to decide if a Project or Research Institute should be linked to one, or to several, or even to all scientific discipline values in which the original scientific discipline was split. Given the same number of entries under such circumstances (~ 50 entries), the CRM team can assess what should be the correct values. Nevertheless, the teams responsible for the follow-up of those entries should always correct the final values.
- ***Issue 2:*** One of the current scientific disciplines in the CRM (Research Infrastructure in all Scientific and Technological fields) does not have a direct mapping under the new classification. Once again, one has to go through all the records with that attribute to understand if a more relevant mapping is adequate. If not, there is also the option to tag such records as "Supporting Activities".
- ***Issue 3:*** The initial import of records in the CRM system was made using Excel (transformed to CSV) information collected by the ESFRI VT. Some of the fields did not have scientific discipline information. As a consequence, a big fraction of records (~570) still do not have scientific discipline information since the responsible persons for following up such records failed to complete this task.
- ***Issue 4:*** The CRM has the possibility for inserting large amounts of data via XLS/CSV files. The major number of options in the new classification categorisation makes these insertions very difficult since it becomes impossible for someone filling an Excel sheet to properly tag the scientific discipline associated to large amounts of records.
- ***Issue 5:*** The actual implementation of the new classification in the CRM user interface may be a problem due to the extensive number of options, and which may compromise the usability of the tool.

Estimated Implementation Effort:

- 15-20 working days for a FTE to implement the new classification until level 3 and proper map the current entries.

Availability of Effort and Priority Level:

- Within the current foreseen workplan, there is no available effort to implement the classification, as the CRM is already over spending due to unforeseen development. However, the availability of effort depends on the priority, which is given by project management not tool managers, therefore is possible if this activity is given a higher priority as opposed to something else. The details would then need to be defined in the final year project planning.

3.4 Operations Portal

Some of the current disciplines with the Operations Portal directly map to the 3rd level in the classification. This therefore places a high priority to implement this level of detail. The integration of 3 levels will require a bit of development effort, which has been outlined below. The initial assumption was to keep only 2 levels into the Portal, which would lead to the potential issues listed below. Other issues are within the structure of the database and technical changes required as well as with VO ID cards, also defined below.

- **Issue 1:** Potential loss of existing information, mostly importantly High-Energy Physics, which would be combined into a more general category of "Physical Sciences". Would VO Managers be happy to be considered in a more global category that the current one? A potential solution could be to implement level 2 and 3, leaving out the generic level 1, but will leave a long list of disciplines making it difficult for presenting statistics.

- **Issue 2:** The current data structure does not allow a VO into 2 or more different disciplines. It means to not only add a new level of information, but completely reviewing the structure in the DB. This part is not possible in the short-term and would require development.
- **Issue 3:** The new classification implies that a significant number of VO ID cards should be changed. This is more than a generic mapping, but would require a long campaign with all VO Managers to ensure that all VO are updated.

Estimated Implementation Effort:

- Around 35 working days for a FTE:
    - Global Tasks
        - Database refactoring + Modification of the classes = 2 days
        - Update of the current and known VO = 5 days
    - VO ID card
        - Integration of a multi level selection = 4 days
        - Modification of the search tool = 5 days
        - Modifications of the work-flow = 5 days
        - Modifications of the interfaces of visualisation = 4 days
    - Metrics and charts
        - Integration of a multi level selection = 1 days
        - Modification of the metrics per discipline = 7 days
        - Modifications of the interfaces = 3 days
        - Modifications of the charts = 3 days

Availability of Effort and Priority Level:

- This effort has not been foreseen for the last year of the project. If this development becomes a priority it means that other tasks should be postponed. The funded effort in PY4 is very limited and some tasks, especially the mini projects, cannot be moved. Moreover, these developments are only meaningful if we have the time to put a campaign in place to update the information. By experience, it will take a significant amount of time to ask VO Managers to update the VO ID cards and change the disciplines. This will need to be evaluated in the final project planning phase.

3.5 Training Marketplace

The Training Marketplace (TMP) currently does not classify its events and materials by discipline although it is something that is high on the TMP priority list. A classification is required so that we can customise the TMP for projects and communities that wish to filter their TMP gadget instance to display events relevant to a scientific discipline. Rather than bring in our own version of a classification we have been waiting on the output of the VT. The eventual use of the new classification within the TMP will only require Level 2, as any further detail is not required or would be useful. Technically there are no issues in implementing the proposed classification in the TMP. The TMP is built on the Drupal framework and will use a pre-built module to implement the changes. Users will be able to multi-select classifications using the standard "control+click" method.

Estimated Implementation Effort:

- Effort was already part of work plan and changes are already being implemented, as it will not require significant effort to make small changes later on (demo ran at EGI CF'13 in Manchester).

Availability of Effort and Priority Level:

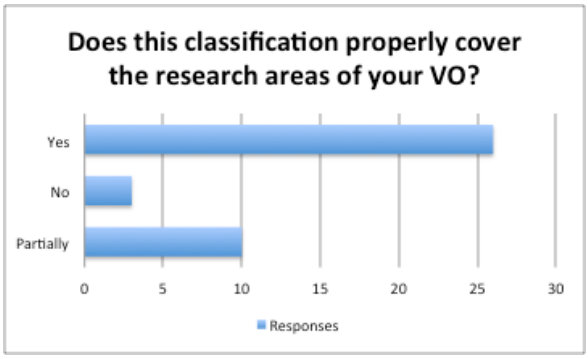
- See above

4 Community Feedback and Public Comments

4.1 VO Manager Survey Feedback

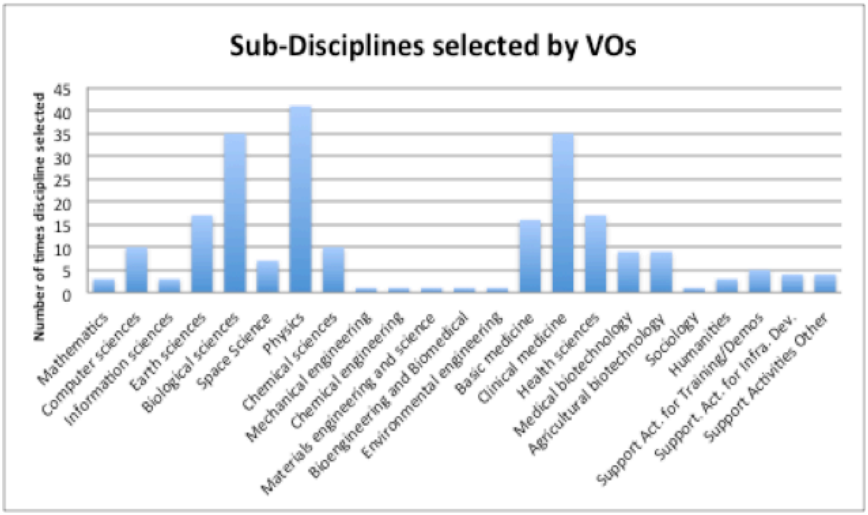
One of the final goals of the VT was to have the new classification validated by VO managers to ensure that their scientific discipline was covered and to offer an opportunity to provide feedback. A very short survey was created and broadcasted to VO managers where 40 VOs responded [R6]. The survey simply asked for the VO name and contact details, to choose any of the scientific disciplines covered by their VO (multi-select of the new classification provided), if the classification properly covered the research areas of the VO (yes, partially, no) and any other comments.

70% of respondents felt that classification properly covered their research areas with a few suggesting that one or two be added (e.g. accelerator physics, seismology). The main issues resulting in the classification were to better define how to incorporate "catch-all" VOs that provide a variety of services and are for any discipline, some unknown. This resulted in part of recommendation 4 to ensure that each tool provides explicit instructions to help users in each tool. Each explanation will instruct catch-all VOs to select all Level 1 categories enabling potential users find these VOs when searching for their specific discipline.



The other issue, which resulted in the only three "no" responses, was not having "Life Sciences" as a specific discipline listed. It was acknowledged by the VT that Life Sciences is a growing community and technology and scientific advances are bridging different fields of science. In fact, it is the very interdisciplinary nature of Life Sciences that makes classifying it very difficult as the number of disciplines it covers are already included in other generic fields of the classification such as Biology, Medical and Health Sciences. This topic was decided to be revisited as more authoritative classifications are defined in the future and moreover, from a user perspective, it is more likely that a researcher searches for their specific field of science such as Biology, Neuroscience, Genetics or another generic field.

One of the best outcomes of the survey was the demonstration of how many scientific disciplines are actually being covered by EGI. This shows that as this new classification is implemented, the level of knowledge and ability to showcase what EGI enables will only increase.



4.2 Public Comments

Taking into consideration the impact of the new classification, it was important to offer an opportunity to the wider community to provide feedback to the classification to help make any final refinements. Therefore, the classification was made available for public comments for a period of thirty days and promoted through a dedicated EGI blog post [R7], inclusion in the monthly NIL dispatch (March) and presented at the Virtual Team Workshop at the EGI Community Forum in Manchester [R8]. Interested parties were able to provide comments either by sending an email to the VT mailing list or by completing an online form [R9].

Two further suggestions were received. One was around linguistics and adding some supporting information, which was modified accordingly, while the other was a suggested tagging system instead of a tree structure. The tagging system was interesting in terms of user flexibility, but had drawbacks in control and maintenance. All issues were discussed in a final VT conference call.

Finally, EGI.eu has been in contact with representatives from the XSEDE project in the US who have expressed interest in this work and are evaluating the classification for potential adoption. Their initial feedback highlighted the need for further work to incorporate their experience and potential for harmonising the classification for common usage. This collaboration will take place over the final year of EGI-InSPIRE and will help shape any future modifications to the classification.

5 Conclusions

There is a clear issue with the current way scientific disciplines are being classified and communicated across EGI. To solve this problem, the Virtual Team brought together a representative from each EGI Tool, engaged with NIL contacts for user community input and followed a step wise process to ensure any new classification would not only cover current and potential usage of EGI, but would also not need to be repeated in the future (at least in the medium-term). Classifying scientific disciplines is a complex issue that has been conducted by a number of organisations and governments with varying outputs. The team made a conscience effort to not duplicate effort by analysing several publicly available classifications while attempting to accurately reflect the research community. Ultimately, the end product of the VT will unify the scientific disciplines making it easier to accurately present usage statistics, harmonise communication both internally and externally, and integrate and account for new communities in the future.

6 Proposed New Classification

6.1 Overview

The proposed classification scales the scientific disciplines from general field of science into functional fields and then further into sub-functional fields. The third level was necessary as many of the disciplines used in EGI are specific in addition to the largest community, High Energy Physics, falling into this category. The VT is fully aware that there are many very specific sub-functional scientific disciplines that could be included in a "Level 4", but it is out of the scope of the VT and not pragmatic in any current or future context within EGI, therefore stops at Level 3.

One aspect that was required to be added to the Frascati FOS classification was "support activities", which was not covered. This category is to include the use of the infrastructure not directly linked to any specific field of science or that spans multiple disciplines or other infrastructure activities. This is not meant to be "multidisciplinary", which is a single VO covering multiple different disciplines. Examples of these have been provided in the table to avoid becoming a catch-all category (e.g. infrastructure development, training). The added functionality of "multi-selection" will provide clarity in not only what fields of science are covered, but also which are in fact "multidisciplinary". X.99 is a reserved code for the "other" category to be used within technical implementation. The following table provides an overview of the first two levels with the third, sub-functional fields, provided further below.

Field of Science	Functional Field
"Level" 1	**"Level" 2**
1. Natural Sciences	**1.1. Mathematics** **1.2. Computer sciences** **1.3. Information sciences** **1.4. Earth sciences** **1.5. Biological sciences** **1.6. Physical sciences** **1.7. Chemical sciences**
2. Engineering and Technology	**2.1. Civil engineering** **2.2. Electrical, electronic and information engineering** **2.3. Mechanical engineering** **2.4. Aerospace engineering** **2.5. Chemical engineering** **2.6. Materials engineering and sciences** **2.7. Bioengineering and Biomedical engineering** **2.8. Environmental engineering** **2.9. Environmental biotechnology** **2.10. Industrial biotechnology** **2.11. Nano-technology**
3. Medical and Health Sciences	**3.1. Basic medicine** **3.2. Clinical medicine** **3.3. Health sciences** **3.4. Medical biotechnology**
4. Agricultural Sciences	**4.1. Agriculture, forestry, and fisheries** **4.2. Animal and dairy sciences** **4.3. Veterinary sciences** **4.4. Agricultural biotechnology**

5. Social Sciences	**5.1. Psychology** **5.2. Economics, finance and business** **5.3. Educational sciences** **5.4. Sociology** **5.5. Law** **5.6. Political sciences** **5.7. Social and economic geography** **5.8. Media and communications**
6. Humanities	**6.1. History and Archaeology** **6.2. Languages and literature** **6.3. Philosophy, ethics and religion** **6.4. Arts**
7. Support Activities	**7.1. Digital Archives** **7.2. Infrastructure Development** **7.3. Training/Demonstrations** **7.99. Other**

6.2 Detailed Classification

The following sections provide the detailed classification that includes the third level of sub-functional fields. Information within ( ) and in *italics* is for additional clarification, not part of the sub-discipline. EGI tools will use the information as part of a "Mouse Over Help Icon" where possible. If not, it will be left out of the user interface.

1. Natural Sciences

1.1. Mathematics

1.1.1. Applied mathematics
1.1.2. Pure mathematics
1.1.3. Statistics and probability
1.1.99. Other

1.2. Computer sciences

1.2.1. Algorithms
1.2.2. Artificial Intelligence *(expert systems, machine learning, robotics)*
1.2.3. Computer architecture
1.2.4. Computer communications
1.2.5. Computer graphics
1.2.6. Computer security and reliability
1.2.7. Data structures
1.2.8. Distributed computing
1.2.9. Human-computer interaction
1.2.10. Operating systems
1.2.11. Parallel computing
1.2.12. Programming languages
1.2.13. Quantum computing
1.2.14. Software engineering
1.2.15. Theory of computation
1.2.99. Other

1.3. Information sciences

1.3.1. Data management
1.3.2. Data mining
1.3.3. Information retrieval
1.3.4. Information management
1.3.5. Knowledge management
1.3.6. Multimedia, hypermedia
1.3.99. Other

1.4. Earth sciences

1.4.1. Atmospheric science
1.4.2. Climate research
1.4.3. Geochemistry
1.4.4. Geology
1.4.5. Geophysics
1.4.6. Hydrology
1.4.7. Mineralogy
1.4.8. Oceanography
1.4.9. Palaeontology
1.4.10. Physical geography
1.4.11. Seismology
1.4.12. Volcanology
1.4.99. Other

1.5. Biological sciences

1.5.1. Aerobiology
1.5.2. Bacteriology
1.5.3. Behavioural sciences biology
1.5.4. Biochemistry and molecular biology
1.5.5. Biodiversity conservation
1.5.6. Bioinformatics
1.5.7. Biological rhythm
1.5.8. Biology
1.5.9. Biophysics
1.5.10. Botany
1.5.11. Cell biology
1.5.12. Computational biology
1.5.13. Cryobiology
1.5.14. Developmental biology
1.5.15. Ecology
1.5.16. Evolutionary biology
1.5.17. Genetics and heredity
1.5.18. Marine and Freshwater biology
1.5.19. Mathematical biology
1.5.20. Microbiology
1.5.21. Mycology
1.5.22. Plant sciences
1.5.23. Reproductive biology
1.5.24. Structural biology
1.5.25. Taxonomy
1.5.26. Theoretical biology
1.5.27. Thermal biology
1.5.28. Virology
1.5.29. Zoology
1.5.99. Other

1.6. Physical sciences

1.6.1. Accelerator physics
1.6.2. Acoustics
1.6.3. Aerosol physics
1.6.4. Astrobiology
1.6.5. Astronomy
1.6.6. Astroparticle physics
1.6.7. Astrophysics
1.6.8. Atomic
1.6.9. Chemical physics
1.6.10. Computational physics
1.6.11. Condensed matter physics
1.6.12. Cryogenics
1.6.13. Fluid Mechanics
1.6.14. Fusion
1.6.15. High energy physics
1.6.16. Mathematical physics
1.6.17. Medical physics
1.6.18. Molecular physics
1.6.19. Nuclear physics
1.6.20. Optics
1.6.21. Particle physics
1.6.22. Physics
1.6.23. Planetary science
1.6.24. Plasma physics
1.6.25. Space science
1.6.26. Quantum physics
1.6.99. Other

1.7. Chemical sciences

1.7.1. Analytical chemistry
1.7.2. Chemistry
1.7.3. Colloid chemistry
1.7.4. Computational chemistry
1.7.5. Electrochemistry
1.7.6. Inorganic and nuclear chemistry
1.7.7. Mathematical chemistry
1.7.8. Organic chemistry
1.7.9. Physical chemistry
1.7.10. Polymer science
1.7.99. Other

2. Engineering and Technology

2.1. Civil Engineering

2.1.1. Architecture engineering
2.1.2. Civil engineering
2.1.3. Civil Protection
2.1.4. Construction/Structural engineering
2.1.5. Transport engineering
2.1.99. Other

2.2. Electrical, electronic and information engineering

2.2.1. Communication engineering and systems
2.2.2. Computer hardware and architecture
2.2.3. Electrical and electronic engineering
2.2.4. Robotics, Automation and Control Systems
2.2.99. Other

2.3. Mechanical engineering

2.3.1. Applied mechanics
2.3.2. Audio engineering
2.3.3. Nuclear related engineering
2.3.4. Reliability analysis
2.3.5. Thermodynamics
2.3.99. Other

2.4. Aerospace engineering

2.4.1. Aeronautical engineering
2.4.2. Astronautical engineering
2.4.99. Other

2.5. Chemical engineering

2.5.1. Chemical engineering *(plants, products)*
2.5.2. Chemical process engineering
2.5.99. Other

2.6. Materials engineering and sciences

2.6.1. Ceramics
2.6.2. Coating and films
2.6.3. Composites
2.6.4. Paper and wood
2.6.5. Textiles
2.6.99. Other

2.7. Bioengineering and Biomedical engineering

2.7.1. Bioengineering
2.7.2. Biomedical engineering
2.7.99. Other

2.8. Environmental engineering

2.8.1. Energy and fuels
2.8.2. Geological engineering
2.8.3. Geotechnics
2.8.4. Ocean engineering
2.8.5. Mining and mineral processing
2.8.6. Petroleum engineering
2.8.7. Remote sensing
2.8.8. Sea vessels
2.8.99. Other

2.9. Environmental biotechnology

2.9.1. Bioremediation
2.9.2. Diagnostic biotechnologies
2.9.99. Other

2.10. Industrial Biotechnology

2.10.1. Bio-derived novel materials
2.10.2. Biocatalysis
2.10.3. Bioderived bulk and fine chemicals
2.10.4. Biofuels
2.10.5. Biomaterials
2.10.6. Bioprocessing technologies
2.10.7. Bioproducts
2.10.8. Fermentation
2.10.99. Other

2.11. Nano-technology

2.11.1. Nano-materials
2.11.2. Nano-processes
2.11.99. Other

3. Medical and Health Sciences

3.1. Basic medicine

3.1.1. Anatomy and morphology
3.1.2. Human genetics
3.1.3. Immunology
3.1.4. Medicinal chemistry
3.1.5. Neurosciences
3.1.6. Pathology
3.1.7. Pharmacology and pharmacy
3.1.8. Physiology
3.1.9. Toxicology
3.1.99. Other

3.2. Clinical medicine

3.2.1. Allergy
3.2.2. Anaesthesiology
3.2.3. Andrology
3.2.4. Cardiac and Cardiovascular systems
3.2.5. Critical care/Emergency medicine
3.2.6. Dentistry, oral surgery/medicine
3.2.7. Dermatology and venereal diseases
3.2.8. Gastroenterology and hepatology
3.2.9. General and internal medicine
3.2.10. Geriatrics and gerontology
3.2.11. Hematology
3.2.12. Integrative and Complementary medicine
3.2.13. Medical imaging
3.2.14. Nuclear medicine
3.2.15. Obstetrics and gynaecology
3.2.16. Oncology
3.2.17. Ophthalmology
3.2.18. Optometry
3.2.19. Orthopaedics
3.2.20. Otorhinolaryngology
3.2.21. Paediatrics
3.2.22. Peripheral vascular disease
3.2.23. Psychiatry
3.2.24. Radiology
3.2.25. Respiratory systems

3.2.26. Rheumatology
3.2.27. Surgery
3.2.28. Transplantation
3.2.29. Urology and nephrology
3.2.99. Other

3.3. Health sciences

3.3.1. Epidemiology
3.3.2. Health care science and services
3.3.3. Health policy and services
3.3.4. Infectious diseases
3.3.5. Medical ethics
3.3.6. Nursing
3.3.7. Nutrition and Dietetics
3.3.8. Occupational health
3.3.9. Parasitology
3.3.10. Public and environmental health
3.3.11. Social biomedical science
3.3.12. Sport and fitness science
3.3.13. Substance abuse
3.3.14. Tropical medicine
3.3.99. Other

3.4. Medical biotechnology

3.4.1. Biomedical devices
3.4.2. Health-related biotechnology
3.4.3. Pharmaceutical biotechnology
3.4.4. Biotechnology and medical ethics
3.4.5. Molecular diagnostics
3.4.6. Biophysical manipulation
3.4.99. Other

4. Agricultural Sciences

4.1. Agriculture, forestry, and fisheries

4.1.1. Agriculture
4.1.2. Agronomy, plant breeding, plant protection
4.1.3. Fishery
4.1.4. Forestry
4.1.5. Horticulture and viticulture
4.1.6. Soil science
4.1.99. Other

4.2. Animal and dairy sciences

4.2.1. Animal science
4.2.2. Dairy science
4.2.3. Husbandry
4.2.4. Pets
4.2.99. Other

4.3. Veterinary sciences

4.3.1. Veterinary anaesthesiology
4.3.2. Veterinary medicine
4.3.3. Veterinary ophthalmology
4.3.4. Veterinary pathobiology
4.3.5. Veterinary radiology
4.3.6. Veterinary reproduction
4.3.7. Veterinary surgery
4.3.99. Other

4.4. Agricultural biotechnology

4.4.1. Biomass feedstock production tech.
4.4.2. Biopharming
4.4.3. Diagnostics
4.4.4. Food biotechnology
4.4.5. GM technology *(crops, livestock)*
4.4.6. Livestock cloning
4.4.7. Marker assisted selection
4.4.99. Other

5. Social Sciences

5.1. Psychology

5.1.1. Biological Psychology
5.1.2. Clinical Psychology

5.1.3. Cognitive Psychology
5.1.4. Comparative Psychology
5.1.5. Developmental Psychology
5.1.6. Educational and School Psychology
5.1.7. Evolutionary Psychology
5.1.8. Industrial–organisational Psychology
5.1.9. Personality Psychology
5.1.10. Positive Psychology
5.1.11. Social Psychology
5.1.99. Other

5.2. Economics, finance and business

5.2.1. Business and Management
5.2.2. Economics and Econometrics
5.2.3. Finance
5.2.4. Industrial relations
5.2.99. Other

5.3. Educational sciences

5.3.1. General Education
5.3.2. Special Education *(learning disabilities)*
5.3.99. Other

5.4. Sociology

5.4.1. Anthropology
5.4.2. Demography
5.4.3. Ethnology
5.4.4. Family studies
5.4.5. Social issues
5.4.6. Social work
5.4.7. Sociology
5.4.8. Women's and gender studies
5.4.99. Other

5.5. Law

5.5.1. Canon Law
5.5.2. Civil Law
5.5.3. Comparative Law
5.5.4. Competition Law
5.5.5. Constitutional Law
5.5.6. Criminal Law
5.5.7. Islamic Law
5.5.8. Jewish Law
5.5.9. Jurisprudence *(Philosophy of Law)*
5.5.99. Other

5.6. Political Sciences

5.6.1. Comparative politics
5.6.2. Empirical data analysis
5.6.3. International relations
5.6.4. Organisation theory
5.6.5. Political economy
5.6.6. Political philosophy
5.6.7. Public administration
5.6.8. Theories of the state
5.6.99. Other

5.7. Social and economic geography

5.7.1. Cultural and economic geography
5.7.2. Transport planning
5.7.3. Urban studies
5.7.99. Other

5.8. Media and communications

5.8.1. Information science - social
5.8.2. Journalism
5.8.3. Library science
5.8.4. Media and socio-cultural communication
5.8.99. Other

6. Humanities

6.1. History and Archaeology

6.1.1. Archaeology

6.1.2. History *(Prehistory; Ancient; Modern world)*

6.1.99. Other

6.2. Languages and literature

6.2.1. General language studies

6.2.2. General literature studies

6.2.3. Linguistics

6.2.4. Literary theory

6.2.5. Specific languages

6.2.6. Specific literatures

6.2.99. Other

6.3. Philosophy, ethics and religion

6.3.1. Ethics

6.3.2. Philosophy of science/technology

6.3.3. Philosophy

6.3.4. Religious studies

6.3.5. Theology

6.3.99. Other

6.4. Arts

6.4.1. Architectural design

6.4.2. Folklore studies

6.4.3. Media Studies *(Film, Radio, TV)*

6.4.4. Musicology

6.4.5. Performing arts studies

6.4.99. Other

7. Support Activities

7.1. Digital Archives

7.2. Infrastructure Development

7.3. Training/Demonstrations

7.99. Other

7 References

R 1	Scientific Discipline Classification VT Wiki page: https://wiki.egi.eu/wiki/VT_Scientific_Discipline_Classification
R 2	Scientific Publications Repository VT Final Report – https://documents.egi.eu/document/1369
R 3	EGI Scientific Discipline Classification (SDC VT Output) – https://wiki.egi.eu/wiki/VT_Scientific_Discipline_Classification_Classification
R 4	Frascati Field of Science and Technologies (FOS) – Org. for Economic Co-operation and Development http://www.oecd.org/science/innovationinsciencetechnologyandindustry/38235147.pdf
R 5	Wikipedia List of Academic Disciplines - http://en.wikipedia.org/wiki/List_of_academic_disciplines
R 6	Survey for VO Manager Feedback - https://www.surveymonkey.com/s/T22H3BB
R 7	EGI Blog Post: New Scientific Discipline Classification for EGI Open for Comments - http://www.egi.eu/blog/2013/04/05/new_scientific_discipline_classification_for_egi_open_for_comments.html
R 8	EGI CF'13 Virtual Team Workshop - https://indico.egi.eu/indico/sessionDisplay.py?sessionId=31&confId=1222#20130410
R 9	New Scientific Discipline Classification for EGI - Open Comments Form https://www.surveymonkey.com/s/KLLV5SS