



HAL
open science

Data-mining and indicators in cyberinfrastructure. Zoom out perspective in French universities cyberspace

David Reymond, Khadija Dib, H. Kromm

► To cite this version:

David Reymond, Khadija Dib, H. Kromm. Data-mining and indicators in cyberinfrastructure. Zoom out perspective in French universities cyberspace. Proceedings of the IEEE EDUCON2012 International Conference: Collaborative Learning & New Pedagogic Approaches in Engineering Education, 2012, Marrakech, Morocco. pp.324–333. sic_01279670

HAL Id: sic_01279670

https://archivesic.ccsd.cnrs.fr/sic_01279670v1

Submitted on 11 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Data-mining and indicators in cyberinfrastructure

Zoom out perspective in French universities cyberspace

Khadija DIB

MIssion Numérique pour
l'Enseignement Supérieur - Direction
Générale pour l'Enseignement
Supérieur et l'Insertion
Professionnelle (MINES/DGESIP)
Ministère de l'Enseignement
Supérieur et de la Recherche
75005 Paris
khadija.dib@education.gouv.fr

David REYMOND
Laboratoire I3M
Université du Sud, Toulon-Var,
83957 La Garde Cedex
david.reymond@univ-tln.fr

Henri KROMM
IMS/LAPS UMR-5218
Université de Bordeaux
351 Cours de la Libération,
33405 Talence Cedex, France
henri.kromm@acthan-expertises.com

Abstract— Aside the e-learning tools, software's, platforms or frameworks, educational organizations provide to their students and staff a variety of digital services through their web portals or cyber infrastructure. Measuring its global usage would be firstly a benefit for policy makers to make decision on objective data, but would provide also useful information to guide on web services tools customization, maintenance or development strategy. We present a non-intrusive, open-source and extensible technical solution that can provide usage indicators on a majority of web digital service environment. The tool can be settled up recursively up in order to provide local, regional or national usage indicators. We will show how the system opens the way to targeted behavioral studies in the context of a performative quality approach of digital services provided by educational organizations and can be mapped to e-learning performance analysis. The examples of indicators creation will be taken from real situations, that is to say in the context of higher education in France.

Keywords-component; Cyber-space, web services, log analysis, usages, Consumer behavior

I. INTRODUCTION

A web site is a very specific communication medium: it offers a variety of functions from the simple presentation of an organization up to specialized digital services. Digital services themselves are various: assistance, audit, internal processes or information chains (such as e-administration or e-education), digital resources access, etc. In the education world these applications and tools are combined to set up the cyber-infrastructure [7], which would include the Virtual Learning Environment [10] (VLE) and have deeply impacted the educational world [9]. We can find in the literature many studies dealing with the evaluation of VLE's, the variety of dimensions to measure [13], splitting qualitative and quantitative aspects, the field of users/students behavior in e-learning dispositive [27], dealing also with e-learning performance shown to be correlated to frequency of usage [22] of e-curricula's, or even studies that zoom out to a wider sense of the VLE [21]: the cyber infrastructure. Many studies again exist about web resources usages in general [18], user's practice of digital services in educational [4] and learning [3] contexts, e-journals [5];[23] or pedagogical resources [8], the

last one tending to explore user's behavior in educational digital context [15]. Much have been written about the cyberspace, showing the changes brought by ICT dispositive in administrative [11] or educational [10] context. It has been shown the importance of uses and users [16] in open education, which can be partially extended to e-learning platform. However there is a lack of studies (or indicators) about the use of the whole set of digital content provided by institutions, defined hereafter by cyber infrastructures: by students or staff, in e-learning or even traditional mode. By measuring the uses not only on the core service (e-learning) but also on peripherals and auxiliary services we expect, as suggested in [1], [21], almost surrounding more efficiently the problem of quality of the global e-learning system. Hence, such a view could help policy makers to develop accurate communication processes, to write curriculums on how to use their e-services adapted to specific non users, or in another perspective to choose the best tool [7] in the very large panel available, corresponding to specific uses. We guess that behavior of students and staffs in cyber infrastructure are intimately closed to their e-learning capability and the need of a global view is hence necessary to address this problem. In this paper, we are going to consider the conditions needed to set up such a system at a national level, which will be based on web log analysis and indicators and extends, as far as we know any traditional web log analysis. A discussion presenting the data-mining capability of the system will show how it could be easily mapped to e-learning analytics. As the context (national) of deployment of this tool is important in the constraints of research and development, we present in this introduction briefly the discipline context, the purpose of setting up broad indicators, and the applicative field of this work: cyber infrastructure in French universities, the definition retained of a cyber-infrastructure, and the need of building a national view.

A. From cyber infrastructure to Web analytics

Cyber infrastructure covers a wide range of applications from the historical one (including office automation) to innovative tools (such as social networking). Using computer network interactions through servers, any user's action can be tracked and stored. The analyses of these interactions are the basis of a wide variety of quantitative studies that opens the

way to targeted communication thanks to users' categorization according to geographic origin, defined interests, but also, in the field of digital services, to the observation of their behavior when using the interface.

From there, metrics are needed to appreciate the usefulness of web digital services. Furthermore, the knowledge of user behavior is fundamental to adapt the different tools - to understand how to develop new paradigms of teaching for instance. The analysis of users' actions and interactions with the services enables to infer certain knowledge of practices and uses of these services [18][6]. More generally, before having recourse to traditional investigation using surveys, the *Web Analytics* models provide useful elements to criticize websites, their design, their organization, or to measure the *Return On Investment* (ROI) and customize, adapt or enhance digital services interfaces according to [19] their users' practice. There is a need for studies dealing with general e-services behavior that would help policy makers with deployment, maintenance or customization, but also to provide knowledge about student behavior when using ICT systems as they are required to do in new ways of teaching [17][25]. As the choice of an e-learning platform is crucial and sometime politically difficult [7], the main background intension at the beginning of the project, was to state quantitatively most appropriate framework for a situation and build up a national best practices over it.

The *Web Analytic Association* (WAA) defines *Web Analytics* as "the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage". This generic definition identifies its usefulness to support policy makers when they choose digital services and define their setting up, maintenance and optimization solutions.

However, this discipline requires the definition of data to be collected, the various processes of analysis and report construction. It consists in aggregating and compiling data to produce simplified and synthetic views. The following step of the analysis attempts to understand the objectives of the users, and possibly to highlight specific findings, identify opportunities and make recommendations for improvement. In order to produce useful comments or recommendations, it is necessary to ensure data reliability. The methods used must be applicable to huge volumes of data. Finally, effective analysis must generate metrics and use *Key Performance Indicators* (KPIs) [18].

The overall analysis involves both the application of a methodology and a strategy to be conducted, leading to the extraction of knowledge from data usage. The report on the data implies the existence of a sub-organization, a policy-making group to be presented to for general strategic objectives.

As e-services are web applications, the essential difference with web analytics are the following:

- We deal with authenticated mode (hence categorization may have to take the internal dimensions of the institution into account while preserving users' privacy);

- Each application can have different modules in several web servers. It can combine various technologies, and each institution may have several e-services...

Keeping this theoretical framework in mind, we will continue by describing successively the field of application and the foundations of the underlying issues to deal with in order to lead to the construction of a suitable method of collection and to show the usefulness of digital services and cyber infrastructure deployment in various layers of control.

Actually, there is a lack of solution to provide uniformed and interoperable logs and even more indicators across the variety of technical solutions. Thinking about it at the scale of large institutions that may also propose different technical tools for the same purpose (multiplicity of e-learning platforms or tools for instance) needs to cross a huge gap in data-mining solution and indicators definitions. In effect, cyberspace offered are usually composed by a range of services from webmail, scholarly services, resources access, e-learning platform, etc., and the architecture (*Single Sign On*, directory services, etc.) to deliver or access to those can be very heterogeneous.

B. *Objective indicators for cyber-infrastructure development*

In this paper, we present an open-source, mostly non-intrusive and reliable system designed to produce rich privacy preserving indicators, with flexible setting up possibilities, which is based on normative frames of reference for the description of users' profiles and on a categorisation of services by kinds of applications.

In the first part of the article, without considering too technical considerations, we will present conditions to set it up, discuss the granularity of user's repository needed and explain the principle of aggregation and indicators construction. In the second part, we will discuss how this dispositive helps to back students in their e-learning curricula, but also help policy makers and ICT managers to select, enhance and customize their cyber infrastructure or e-learning system. The discussions will be supported by results coming from a pilot deployment from 2009 to 2011 in several French universities. An abstract generalization using our research paradigm in usage indicators definition will also show that the interoperability offered by the system allow interuniversity's comparisons that can be used to infer user's behavior avoiding technical or local/regional biases.

The problem addressed here will be presented in the context of e-services in the French academic virtual space, and extended to a more generic consideration due to the needs of the project: the need for policy makers to set usage indicators of large bunch of web applications, helping them in their governance objective at different scales. As the system described is open-source, flexible and very extensible, it opens straightforward to behavior analysis avoiding well-known biases due to technical reasons. In a technical point of view, the system presented is a new and simple way to produce usage indicators in authenticated and centralized environment dealing with many web applications. Before presenting the dispositive, we present the context of development, expressing the

motivations and the constraints shared by many high educational organizations [7].

C. Development of French faculties cyber infrastructures

Historically, the management of technology department (MINES) within the French Ministry of National Education, Higher Education and Research (MESR) has initiated several actions to develop e-services and their uses in higher education institutions.

The service launched in 2003 a call for projects UNR. Objectives tripartite contracts are signed between consortiums of higher education institutions, regional councils concerned, and the MESR.

These contracts included several purposes:

- The development by institutions digital workspaces or cyber infrastructures destined to the university community.
- Providing access for all students to these services (individual and group equipment's, networks, individual access to broadband, etc.).
- The supports to develop the uses of these services.

Today, the MINES monitor and provide national coordination the 17 UNRs covering the country (see Figure 1.). A steering committee, comprising the various bodies in relation to the project meets regularly.



Figure 1. The UNR's in the French territory

The web digital services deployed within the UNR revolve primarily around [32]: education, pedagogy, communication, documentation, office, university life and relationships with companies. But many other original web services such as podcasts or social applications enrich locally the bouquet of digital services.

Currently 98% of French students have access to a bunch of digital services. In this context, the MINES wishes to provide institutions with a technical device enabling them to track the usage of those deployed and to produce multi-level dashboards to assist policy makers and managers in digital resources and information system in general. We will explain the solution

develop hereafter and show how this primary objective can be enlarged using the data-mining capability to e-learning research questions. We will first present hereafter the definition used for cyber infrastructure in this paper.

D. Faculty cyber infrastructure definition

The notion of digital service can be handled differently [10][11][13]. Following an investigation initiated at the national level with the universities IT managers in charge of deploying digital services; we collected in 2009 pragmatic views that lead to our definition. We will use the following technical definition, based on our survey [28]. A cyber infrastructure is defined in three points:

- The applications or services are in client and web server mode (the users have to use a web browser embedded - or not - within client software), from any device (PC, station or mobiles...).
- The access to the application requires authentication via a centralized authentication service (Single Sign On - SSO) based on an identification directory.
- Identification is performed using a centralized directory of members of the organization or even a federated directory.

Hence, the whole bunch of services deployed by an organization (or group of organizations) that meet the previous statements creates a cyber-infrastructure.

Thanks to this definition we are able to deal with all the digital services of an institution belonging to a web access mode via authentication. However, in practice, the architecture can follow different patterns (e.g. it is made up only with an educational framework, i.e. the Moodle framework, or JASIG u-portal). This does not affect the previous definition: the framework authentication sub process can be considered as the "SSO" of its own and will provide web application to users, just like another web application. The unique condition in this case would be to access to each application with almost a singular URL.

E. National view

The need to produce quantitative indicators of usage at this level is straightforward [16] almost if one thinks in best-practices elaboration. The previous definition used permits circumventing perimeters of cyber infrastructures. Historically, as the faculties were free to set their own architecture, making their own choice in technologies or application deployed, the result today is that cyber infrastructures are very different from one to other.

Basically, three problems are addressed here to meet the objective of considering web application usage and behavior at the expected scales:

- 1) The need for a dispositive to be non-intrusive and mostly technically independent from the very heterogeneous solutions provided around the territory;
- 2) The need to produce indicators that allows comparing usage on very heterogeneous bunch of applications;

3) The need to guarantee data consistency.

This application AGIMUS (*Application de Gestion d'Indicateurs de Mesure des Utilisations de Services*), developed under CECILL license (mostly like a French GPL, see <http://www.cecill.info>), will threat front-end web-logs of the digital services deployed in the institution and construct multiple scale indicators, using a recursive deployment. Before presenting the solution we address the two last points using traditional librarian and information science technique: nomenclature for consistency and categorization for summarization. Basically, this will provide a mode to compare easily rough usage statistics on several technologies.

The way to deal with the complexity inherent to the variety of elements is to use indicators that are built using a categorization of services, sort of meta-indicators.

II. AGIMUS' DESCRIPTION

Following a call for tenders issued in October 2009, the AGIMUS application was developed is now functional [31]. The proposed solution meets the technical difficulties of constructing indicators that are adapted to the variety of categories and application architectures used to provide e-services and it enables institutions to include the context of their deployment. Furthermore, it offers the possibility for an institution to create its own indicators. The same device is used to produce aggregates and building trees of data collectors to measure the usage of services and it is flexible enough to be used across the landscape of the Higher Education field in

France (institutes, engineering schools, universities, etc. are as many paradigms for sharing resources or objectives which may require indicators to measure the use of digital services). The usefulness of the device reveals itself at the micro level – the level of the service managers with specific activity indicators, at the macro level – for the construction across the territory of aggregate indicators and comparable measures of use - but also in the middle, at the level of UNRs to drive the deployment of their services. We will give a more detailed account of the key points of this application in the next part: data collection method, construction of indicators, and data enrichment for aggregation and meta-indicators [30]. The openness of the application enables a functional range to evolve dynamically according to the way institutions and managers want to use them.

A. General function

The application AGIMUS produces indicators on the use of web applications (identified by URL). Coming from user's interactions with the cyber infrastructure, log data is processed to calculate segmented indicators by user's category and of course in time dimension. A data-mining console offers the indicator scorecards

Once established in a set of faculties, the device is recursively set up a structure (say "UNR" here but any kind of hierarchical organizational system can be built up), to aggregate the data provided from the faculties instances and build inter-institutions indicators (UNR indicators) and, by the same process, between UNRs (Ministry indicators). Using the

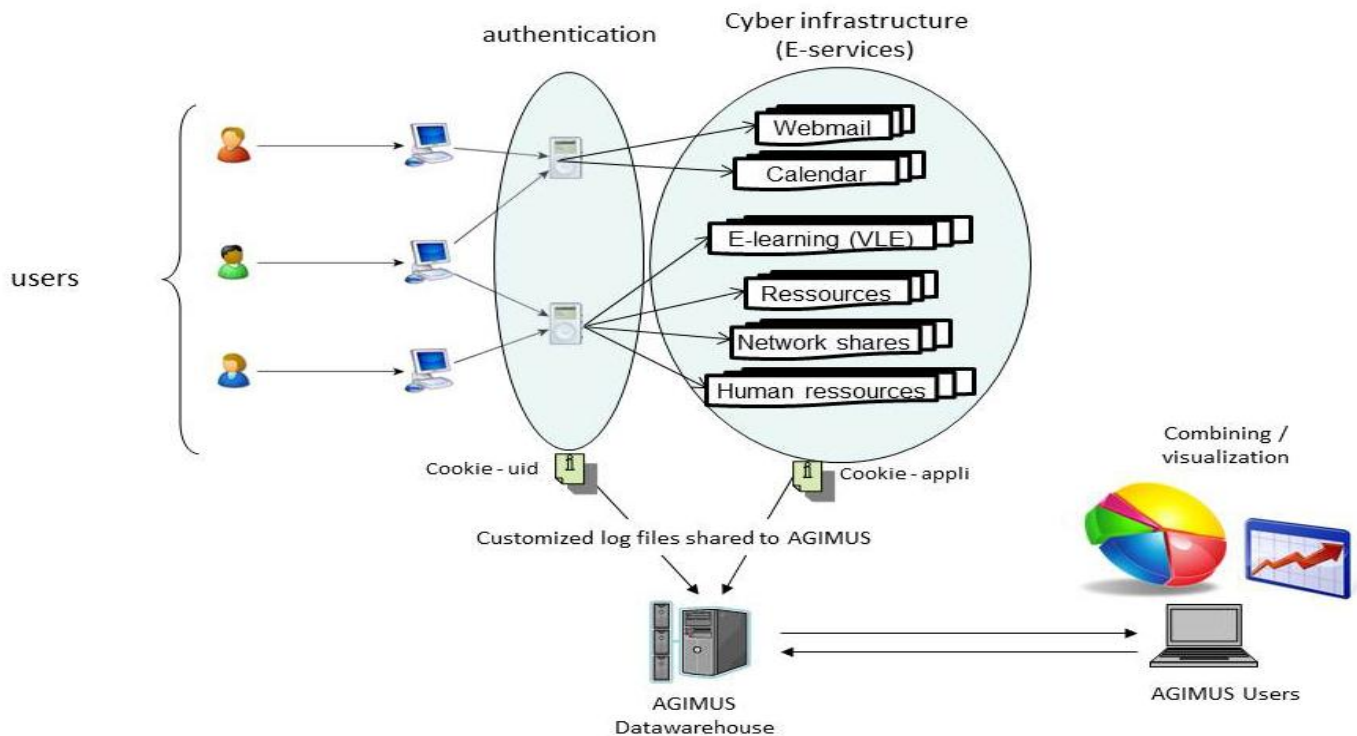


Figure 2. AGIMUS technical principles, weakly intrusive in wide range cyberspace technologies

categories paradigm, the provided indicators are technology-independent from the tracked applications services.

B. The application AGIMUS

AGIMUS (actually in version 1.4) is a web application installed on a server providing PHP 5.2 interpretation capability and connected to a MySQL server to host the data warehouse.

C. A non-intrusive deployment

1) A web application

Figure 2 shows general scheme of AGIMUS installation into a cyber-infrastructure. AGIMUS receives customized data logs marked by cookies and coming from front-end applications and from the authentication system. At the actual stage of development, *Apache* (<http://www.apache.org>) and *lighttpd* (<http://www.lighttpd.net/>) web servers are correctly supported with the Central Authentication Service (CAS) of the *Jasig community* (<http://www.jasig.org/cas>) in its latest versions.

The logs are provided to AGIMUS via classical network sharing or logging system. AGIMUS will hence detect the cookies in the log files produced (one per front-end server). This phase will drastically reduce quantity of log data and acts

Thanks to inner users repositories, users IDs are retrieved and immediately replaced by an anonymous number but their description is enriched with characteristic elements whose semantics is in the field of academic organizations (course, level, discipline, department...). Staff user characterization does not use such granularity, only affiliation is used. In order to be aggregated correctly, the student user profile description has to be based on a nomenclature. That is why we have chosen the *eduPerson* entry in the *Supann* scheme, normally used in LDAP directories as a recommendation for our institutions (<http://www.cru.fr/activites/supann2/>)

Thanks to these dimensions AGIMUS will offer the possibility to produce a wide range of indicators and their variations on user's dimensions. According on governance objectives these indicators aim at providing objective information on questions such as:

- Which students are frequently users of the VLE?
- What percentage of students of my institution are web scholarship service users?
- What percentage of students in philosophy uses the webmail service?
- What is the penetration of the entire range of the cyber

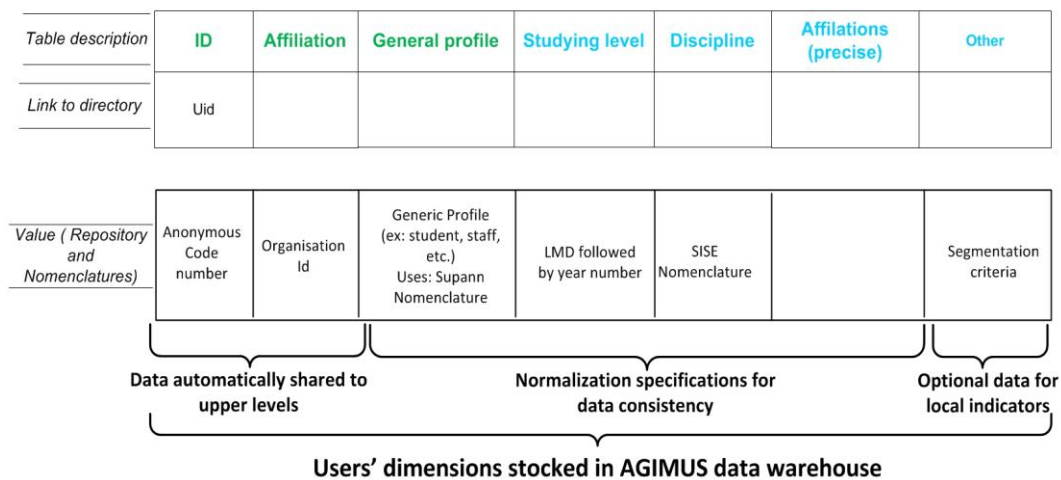


Figure 3. The Faculty users descriptions stocked in AGIMUS data warehouse

as a filtering process. Data are processed daily by the system that enters at this time a special phase detailed hereafter.

2) User's description anonymization and enrichment processing

While AGIMUS filters logs from front-end server, it will use the cookie in the SSO log line to enrich and anonymize the users' description. As well as data must be share out from the organizations, nomenclatures are needed here to guarantee data consistency. Figure 3 describes the tables in the data warehouse used for user's description; we distinguish here the columns data shared automatically, columns needing normalization specifications and optional entries.

infrastructure in the population of lecturers/researchers?

- Do students from hard sciences departments use the cyber infrastructure more frequently than students in social science and humanities?

For the biggest population of our institutions (students), this initial segmentation, is needed to improve the indicators and adapt them to natural groups (level of education, discipline, field and graduate enrollment).

Student description will follow the standardized SISE nomenclature to allow the calculation of metrics at the inter-institution levels (N +1) or higher. This concerns the upper

stage: connecting recursively several AGIMUS repositories to build up inter-cyber infrastructures indicators.

3) Virtual hierarchical recursive setup and aggregation method

As shown in Figure 4, a recursive set up process is used at the UNR level (or hierarchical virtual level settled up) to aggregate data coming from faculties instances.

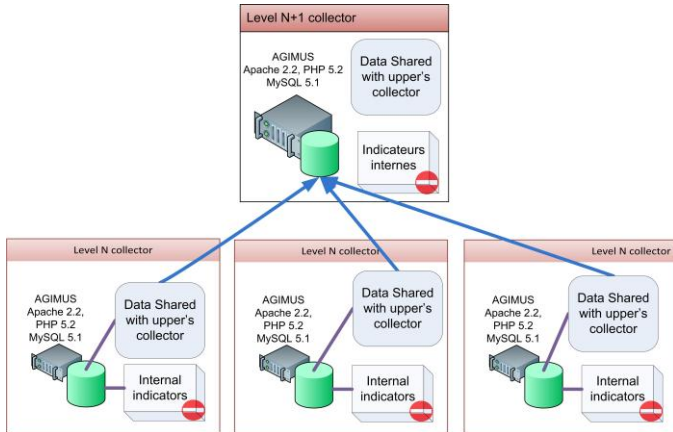


Figure 4. Generic aggregation scheme of AGIMUS collectors

This process can be repeated as needed, in order to aggregate independent monitored services. Thanks to this architecture, each level can gather and construct its own and internal indicators, and deposit ruled controlled data to a (virtual) superior hierarchical instance.

AGIMUS dispositive allows to aggregate data coming from multiple cyber-infrastructure to compare practices and draw an appropriate benchmark. The aggregation operation consists of a combination of entities according to predefined criteria in order to facilitate understanding and interpretation. Kromm [18] defines a logical aggregation as a three-step process: partition, instantiation and reduction.

- Partition is the combination of different entities according to homogeneous or heterogeneous grouping criteria (in our discussion, the faculties' cyber infrastructure). We construct homogeneous partitions clusters thanks to the nomenclature. We call this the categorization stage.
- The second step consists in associating one or more parameters to the object: for example, the "profile" object is an instance of the "User" attribute and is created by counting the number of logs corresponding to it.
- Hence, in the reduction step, AGIMUS generate aggregated data from detailed data: each level of aggregation then manipulates the data which are classes of information handled at lower levels of decomposition.

Based on these principles, our system enables to define the most relevant type of reduction according to the data handled and grouping/partition made.

4) Indicator setting objectives

We invite the interested reader for a deeper discussion [31] in AGIMUS indicators setting cautions [29], as these indicators come with a global multi-level [30], recommendation plan resumed hereafter briefly:

- Creating a global strategy quality plan for cyber infrastructure development
- According to AGIMUS indicators
- Define policy makers actions

Some recommendations on how to create indicators that will correspond to AGIMUS, must be followed to address this general problem at a national level.

A first experiment on the scale of a UNR was conducted in 2008 in Aquitaine [28]. This experimental work, using a non-free technology, was meant to show the relevance of quantitative usage indicators in decision making at every level and also to determine its operational feasibility [29]. However, it shows its limits due to its intrusiveness inside the e-services programming code (very complex to implement sometimes). That is why we have preferred to develop a non-intrusive, free and easily scalable solution, adaptable to the huge range of e-services.

D. AGIMUS' indicators

AGIMUS comes with already settled basic indicators. They remain into three categories, conforming to the educational hierarchy level as described by Jacquinot and Fischer [17] (p. 20-21) according to their use objectives: monitoring and maintenance indicators, management indicators and operating indicators. We describe them hereafter, inviting the reader to keep in mind the possible variation in the user's dimensions.

1) Monitoring Indicators

TABLE I. lists the indicators produced by the device at the N and N+1 levels.

TABLE I. PREDEFINED MONITORING INDICATORS

E-service operator (N level)	Services aggregate (N+1)
Number of uses	Number of uses
Effective daily users	Effective daily users
Number of operational services	

Monitoring indicators provide information about the daily operational services and their degree of use. They are expected to be used in a management plan: planning an update of applications, allowing low user incident modifications in choosing off-peak periods, etc. AGIMUS acts at this level as many other systems by centralization of activity logs on the system, but also allows refining the choices by splitting activity information on user's categories. At the N+1 level, these indicators give a detailed comparative view of usage in institutions at N level. They can be used to estimate the degree of usage of each service and compare them. In the upper levels, the analysis should generate a best-practice perspective (for instance regional), and exchange between managers methods or products that seems to be best suited for students and staff.

2) Governance Indicators

TABLE II. presents the governance indicators describing the services and categories of service to outline the average behavior of the users. Behavior is simply expressed in terms of the duration of a session and the frequency of connection. Whether they are produced for a single institution or by aggregation, these indicators provide information about the effective use of services. The institution can use the raw connection rate to complete the analysis of usages.

TABLE II. PREDEFINED GOVERNANCE INDICATORS FOR VIRTUAL LEVELS N AND N+1

Digital service operator (level N)	Aggregated level (N+1) Category view
Digital service covering	Categorized digital service covering
Average session duration	Average session duration
Average frequency of use	Average frequency of use
Raw connecting rates	Raw connecting rates
Types of browsers	

4.4.3 Operating Indicators

TABLE III. describes the operating indicators specific to the AGIMUS device. These indicators are designed to assist in its implementation and the validation of the audit process of digital services. At N+1 level the data is reconstructed by a summative aggregation.

TABLE III. OPERATING INDICATORS OF AGIMUS

Indicators	Definition
URL detected	Allows the identification of new applications Enables to refine granularity (for instance creates the 'mail sent' indicator)
Digital audited services	List of web applications declared and audited
Population numbering	The number of the different population is used when calculating rates.

3) Life cycle operations

The use of these indicators is based on the principles of control and governance, derived from enterprise modeling and the GRAI method. We refer to its original definition Doumeingts [11] and [34] for a widest presentation.

To each type of indicator corresponds firstly a frequency of update and secondly a delay to complete the action plans (all the decisions taken at the end of the result analysis). Consequently, the various indicators discussed above are aimed to be used temporarily in recursive steps of update and use:

The governance Indicators are expected to be use at a long-term level, while Monitoring indicators are expected to be used at an average term. Operating indicators should be used at a short term level.

III. FIRST RESULTS FROM PILOT TEST

We provide here some examples of graphics directly coming from AGIMUS instances deployed in partners faculties (*Université de Pau et des Pays de l'Adour* – UPPA- and *Université de Tours*). These faculties were pilots, and the success of the deployment makes the Ministry engaging a national deployment in 2012.

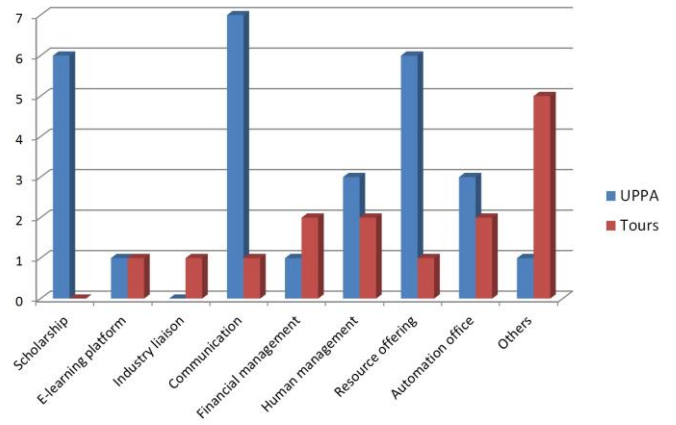


Figure 5. Variety of web application deployed in the cyber infrastructure of the UPPA (2010) and the university of Tours (2011)

Figure 5. shows the composition of the cyber infrastructure of the UPPA (in 2010) and the *Université de Tours* (2011). We use application category instead of namely each application to figure them in the graphic. Each dimension of the graphics shows a category (Scholarship, e-learning platform, Industry liaison, etc.) and the number of application attached to it. For instance, this cyber infrastructure of UPPA deploys six resource offering (*documentation*) applications and 1 for Tours. The category “other” will be used to check novelties at National level. The categories aim to compare avoiding the technologies used besides. Such graphics are used in institutional partnership collaboration (funding’s research) to describe easily what is done at a time. Determining each service rate use comes as evidence to target objectives and help policy management. Such information is expected to become a basic quantitative pre-study to precise and qualitative behavior studies to perform the cyber infrastructure. This comes as an example of the variety of cyber infrastructures in high education organizations. At the UNR level, such a view avoids unfair comparison: usages of both cyber infrastructures must be related to comparable number of application in the categories; hence the comparison of rate of use can offer useful information in behavior on the specific applications. At this level, AGIMUS offers starting point for best practices analysis. Of course, more precise dashboards are needed to construct a precise analysis that would be beside the objectives of this paper. We provide them as suggestive options in statistical behavior analysis and/or policy maker knowledge building process helping them in information system management task. Data-mining capability, flexibility and other options of AGIMUS will be describe after but must be kept in mind to estimate the range of applications of the system.

IV. A NEW PATH TO E-LEARNING ANALYTICS

As introduced before, AGIMUS offers the capability of using data-mining techniques like clustering [14] in order to reveal generic behavior on cyber-infrastructure. Figure 8 shows the evolution of access of two categories of users in the time of the e-learning platform: students (*étudiants*) and teachers (*enseignants*). We can remark the peaks of teacher’s uses before student’s peaks. As this data comes at the beginning (the

first two weeks) of school time, no further conclusion is possible.

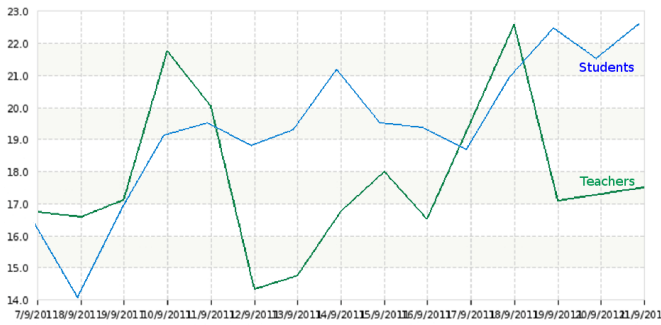


Figure 6. 2 weeks of time evolution of student connections and teacher connection

But one can easily imagine combining such information with other information, such as learning analytics data, or comparing to others usages in partner's faculties or in the UNR to perform the cyber infrastructure. Figure 7. shows the main dimensions of the data warehouse opening to multiple selections (by SQL requests) to provide objective data on targeted questions. As AGIMUS offers the possibility to insert particular flow of data (coming for instance from e-learning platform). Hence, the student dimension can be augmented with e-learning specific data (such as number of assessments, time spent in course, discussion activity in forums, number of attempts, etc.) to develop more specific indicator of e-learning performance. At the time of writing this paper, we could not

provide more precise example, gathering sensible data from several institution is not easy but we hope that the system will open the community to dress up deeper analysis to analyses the dependencies between discipline and e-learning performance, or cyber-infrastructure skills and e-learning capability, etc. as it provides a straightforward way to enter the quantitative part of such analysis.

V. CONCLUSION

In the case of the French universities we plan to set-up the deployment of AGIMUS into a majority of cyber infrastructures, and construct regional monitoring system (UNR levels) and, so on, the Ministry level in 2012. The objective of the MINES is to offer quality service to the end user, by providing Faculties a system that automatically creates quantitative indicators about the use of the cyber infrastructures of the institutions. The immediate purpose of these indicators is to help policy makers to enhance these digital services and their usage.

The challenge addressed is to participate in decision-making improvements, changes and adaptations that will promote the adoption of the cyber infrastructure (including e-learning system) by users, but also to discuss their effectiveness and adapt the web application to their uses. On the other hand, the objective of such a challenge is also to capitalize knowledge of actual usage transversely (between the different structures, even competitive ones), but also vertically (through an aggregation of different indicators from institutions, UNRs or other structures, and finally, at a national

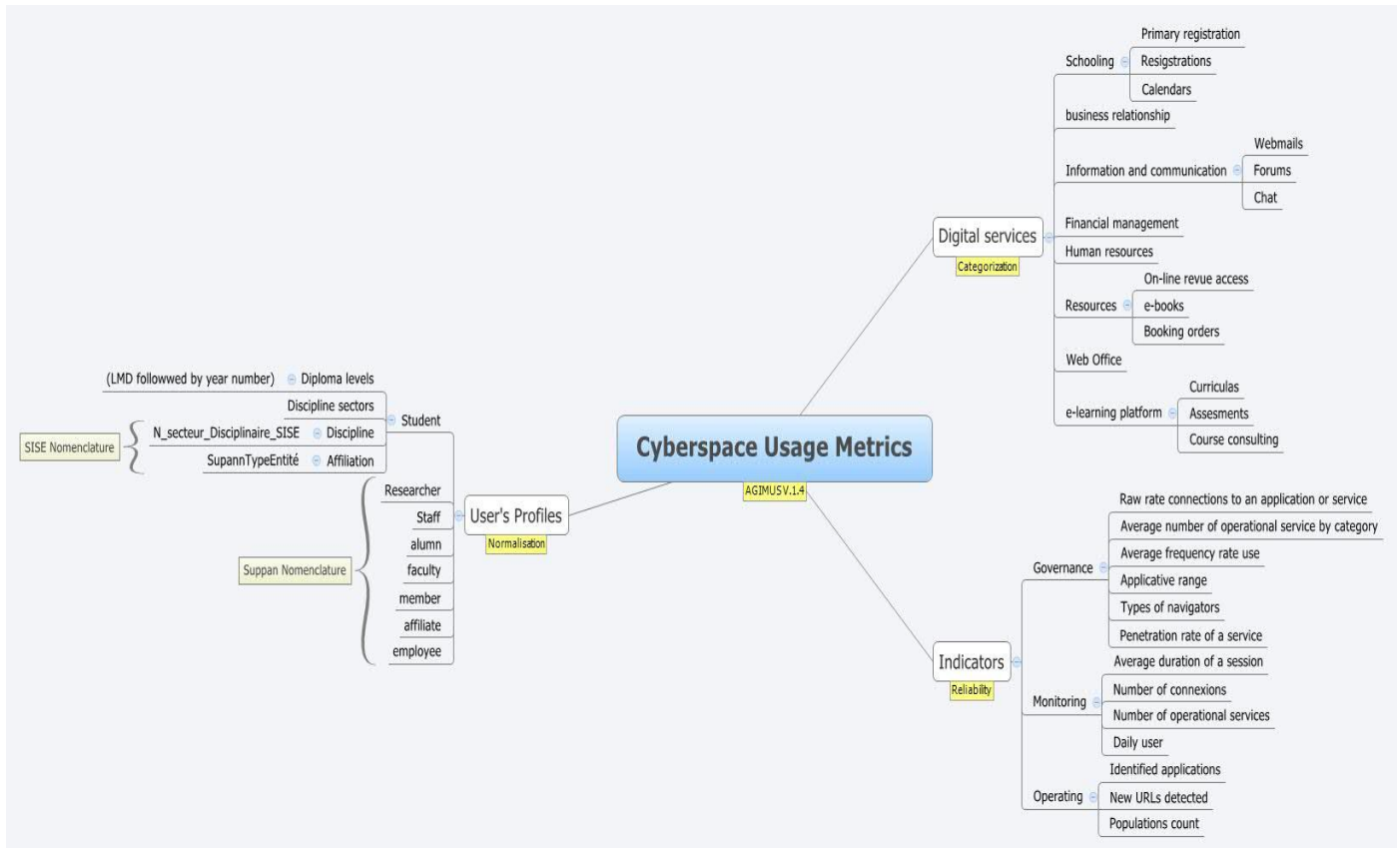


Figure 7. The two main dimensions in AGIMUS datawarehouse (users and digital services), and the list of pre-built indicators

level, from the ministry). These aggregations would provide a useful data source in the construction of "good practices" that would help to reduce the costs of experimentations, service offering and help setting up improvements by sharing them regionally or nationally so as to be competitive with similar services provided out of our institutions.

AGIMUS, with its modular implementation, is non-intrusive and adaptable to a wide variety of cyber infrastructure or isolated e-services. The extension presented, using the URI granularity, opens the consideration of an e-service as a combination of several web applications and is paving the way to set up specific (and missing) quantitative indicators in complex systems such as e-learning frameworks [13] or even resources offering services [3][4] as expected in general sense [1]. The method of production of the indicators is associated to a method of control underpinned by research currently very active internationally [24]. The granularity of the indicators based on natural groupings of academic profiles, and a standardized nomenclature for their description that offers the possibility of an active piloting process [20], combined with the maintenance and the deployment process of digital services [7].

By offering usage indicator respecting privacy by according to separate population group in the organizations maintaining the cyber infrastructure, AGIMUS is a way to understand user behavior using web application. The principle of AGIMUS is to produce indicators of statistical behavior and anonymous digital services paving the way for qualitative behavioral studies at a lower cost but also by circumventing classical biases in those studies as local or regional biases but also biases coming out from the technology of the systems as experimental ways. Furthermore, knowledge provided at the student granularity scale allows deploying corrective policies such as communication or teaching on cyber-infrastructure usage that can be dedicated specially for those who would need this information. Targeted [19] behavioral qualitative analyses are hence more easily designed thanks to the quantitative behavior statistics gathered. We expect providing, as soon as the first deployments are done, a deeper analysis showing the relations between ICT skills and e-learning performance and provide to the community large scale usage and behavior data.

6. ACKNOWLEDGMENTS

The authors would like to thank the many contributors to this development. In particular Nicolas CAN (UPPA) and Julien MARCHAL (*Université de Nancy*) for their precious technical advises and support.

REFERENCES

- [1] Abdullah, F. (2005) "HEdPERF versus SERVPERF: The quest for ideal measuring instrument of service quality in higher education". *Quality Assurance in Education*, Vol. 13, No. 4, pp 305-327
- [2] Badolato, A.-M., M. Colin, P. Houdry, S. Launay and D. Lechaudel. 2009, « Ressources électroniques des portails de l'inist : analyse qualitative par comparaison des consultations et des facteurs d'impact », dans *Actes du colloque VSST'2009*, Nancy.
- [3] Barral, S. 2007, « Indicateurs d'usages des ressources électroniques », technical report.
http://www.sup.adc.education.fr/Bib/Acti/Electro/mission_barral.pdf
- [4] Boukacem, C. and Shöpfel, J. « Statistiques d'utilisation des ressources électroniques : le projet counter », *Bulletin des Bibliothèques de France*, vol. 50, N°4, 2005, p. 62–64.
- [5] Boukacem, C. and Shöpfel, J. "On the usage of e-journals in French universities". *The Journal for the Serials Community*, Volume 21, Number 2 / July 2008, pages 121 – 126.
- [6] Boukacem-Zeghmouri, C., (éd.). 2010, *L'information scientifique et technique dans l'univers numérique : mesures et usages*, ADBS éditions, 319 p.
- [7] Brown, M, Paewai, S and Suddaby, G. "The VLE as a Trojan Mouse: Policy, Politics and Pragmatism" *Electronic Journal of e-Learning Volume 8 Issue 2*, 2010, pp 63–72, available online at www.ejel.org
- [8] Colin, M. and D. Lechaudel. 2010, « Connaissance des consultations des ressources électroniques du CNRS : méthodologie et applications », in [6], p. 129–140.
- [9] Craig, A., Goold, A., Coldwell, J., Mustard, J. (2008) "Perceptions of Roles and Responsibilities in online Learning: A Case Study", *Interdisciplinary Journal of E-Learning and Learning Object*, Vol.4, 205-223.
- [10] Crane, Gregory, Brent, Seales, and Melissa Terras. 2009. Cyberinfrastructure for Classical Philology. *Digital Humanities Quarterly* 3, no. 1; Special issue: Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure.
<http://www.digitalhumanities.org/dhq/vol/3/1/000023.html>
- [11] Doumeingts G., *La méthode GRAI*, Thèse d'état, Université de Bordeaux I, 1984.
- [12] Gerbod, D. and Paquet. F. *Les clefs de l'e-administration*, Pratiques d'Entreprises, Editions Management et Société, 2001
- [13] Dyson, M., Barreto Campello S., 'Evaluating Virtual Learning Environment'. *The Electronic Journal of e-Learning (EJEL)*, p. Vol 1, N°1; Feb 2003, p. 11–20, available online at www.ejel.org
- [14] Han J., Kamber M., and Pei J., *Data Mining: Concepts and Techniques*, 3rd edition, Morgan Kaufmann, 744 p., 2011
- [15] Harley D., "Use and Users of Digital Resources: A survey explored scholar's attitudes about educational technology environments in the humanities". *EDUCAUSE Quarterly* N°4, 2007
- [16] Harley, D. "Why Understanding the Use and Users of Open Education Matters". In *Opening Up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge*, edited by Toru Iiyoshi and M.S. Vijay Kumar. Cambridge, MA: MIT Press, 2008.
- [17] Jacquinet G. and Fichez E., *L'université et les TIC*, Collection: Perspectives en éducation et formation, De Boeck Université, 2008. 328 pages.
- [18] Jansen, B. J. 2009, "Understanding User-Web Interactions via Web Analytics", N°6 in *Synthesis Lectures on Information Concepts Retrieval and Services*, Morgan & Claypool Publishers.
- [19] Khan, R., Lewis, M. and Singh, Vishal. 2008. "Dynamic Customer Management and the Value of One-to-One Marketing". *Marketing Science*, Vol. 28, N°6, 17 pages, pp 1063-1079
- [20] Kromm, H., *Contribution à une méthodologie d'analyse de la cohérence entre les objectifs de conception et d'exploitation d'un système de production*, PhD thesis, université Bordeaux I, 2002
- [21] Martínez-Argüelles, M, Castán, J and Juan, A. (2010) "How do Students Measure Service Quality in e-Learning? A Case Study Regarding an Internet-Based University" *Electronic Journal of e-Learning Volume 8 Issue 2 2010*, (pp151 - 160), available online at www.ejel.org
- [22] Nakayama, M and Yamamoto, H. "Assessing Student Transitions in an Online Learning Environment" *The Electronic Journal of e-Learning Volume 9 Issue 1 2011*, (pp75-86), available online at www.ejel.org.
- [23] Nicholas, D, Huntington, P, "Evaluating the Use and Users of Digital Journal Libraries". In: *Digital Libraries*, Ed. Papy, F, 2008, London: ISTE Wiley
- [24] Nicholas D, Rowlands I. "Digital consumers: case study virtual scholars. A deep log analysis". In *L'information scientifique et technique dans l'univers numérique : mesures et usages*. ADBS, 2010, pp 27-42

- [25] Paivandi, S. « L'enseignement à distance : un facteur de changement à l'université », in [33], chap. 13, p177-188
- [26] Pinède, N. 2010, « Rôle stratégique des sites web pour la gouvernance des organisations universitaires », in Session poster du colloque international "l'Université à l'ère du numérique" (CIUEN 10), Strasbourg.
- [27] Poll, R. 2007, « Benchmarking with quality indicators: national projects », *Performance Measurement and Metrics*, vol. 8, N°1, p. 41–53.
- [28] Reymond, D. and Dib, K. 2008, « Indicateurs d'usage des services numériques déployés au sein des universités numériques en région », dans *Actes du colloque international "l'Université à l'ère du numérique"* (CIUEN 08), Bordeaux.
- [29] Reymond, D. and Dib, K., 2009, « Vers une interopérabilité de la mesure d'usage des ENT : enjeux, objectifs et méthode », *Intelligence collective et organisation des connaissances*, in *actes du 7e Colloque du Chapitre français de l'International Society for Knowledge Organization (ISKO)*, M. Hassoun et M. El-Hachani (Eds), Ensib, Lyon, p. 287–293.
- [30] Reymond, D. and Dib, K., 2010a, « Mesure d'usage et organisations multi-échelles : indicateurs et méta indicateurs d'usage des ENT. », dans *colloque international « l'Université à l'ère du numérique » (CIUEN) juin 2010*, Strasbourg
- [31] Reymond, D. and Dib, K. 2010b, « Mesures d'utilisation : vers un pilotage intelligent des services numériques de l'enseignement supérieur », In *Actes du Colloque International AFRICAMPUS, Les usages intelligents des technologies de l'information et de la communication dans la réorganisation universitaire*, Dakar Sénégal, to appear.
- [32] SDET, 2006, « Schéma directeur des espaces numériques de travail (SDET) », Ministère de l'enseignement Supérieur et de la Recherche, version 2.0
ftp://trf.education.gouv.fr/pub/educnet/chrgt/sdet/SDET_v2.0.pdf.
- [33] Sun-Mi K. and Verrier C., Le plaisir d'apprendre en ligne à l'université. Implication et pédagogie, Bruxelles, De Boeck Université « *Perspectives en éducation et formation* », 2009, 228 pages.
- [34] Vallespir B., Ducq Y., Chen D., Enterprise modeling for interoperability, *16th IFAC World Congress*, ISBN : 0-08-045108-X, Prague, July 4-8 2005