ESO's User Portal: Lessons Learned

A. M. Chavan[1], L. E. Tacconi-Garman, M. Peron, F. Sogni, D. Dorigo,
P. Nass, N. Fourniol, D. Sforna, K. Haggouchi, M. Dolensky;
European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany

ABSTRACT

ESO introduced a User Portal for its scientific services in November 2007. Registered users have a central entry point for the Observatory's offerings, the extent of which depends on the users' roles -- see [1]. The project faced and overcame a number of challenging hurdles between inception and deployment, and ESO learned a number of useful lessons along the way. The most significant challenges were not only technical in nature; organization and coordination issues took a significant toll as well. We also indicate the project's roadmap for the future.

Keywords: Portals

1. INTRODUCTION

The first document describing what was to become ESO's User Portal[2] was released on December 29th, 2004 (see [2] for the description of a project based on those initial ideas); the Portal itself went live on November 15th, 2007. Now all of ESO's science-related, Web-based services are accessible from a single, secured access point; all users are uniquely identified and can manage their own profiles.

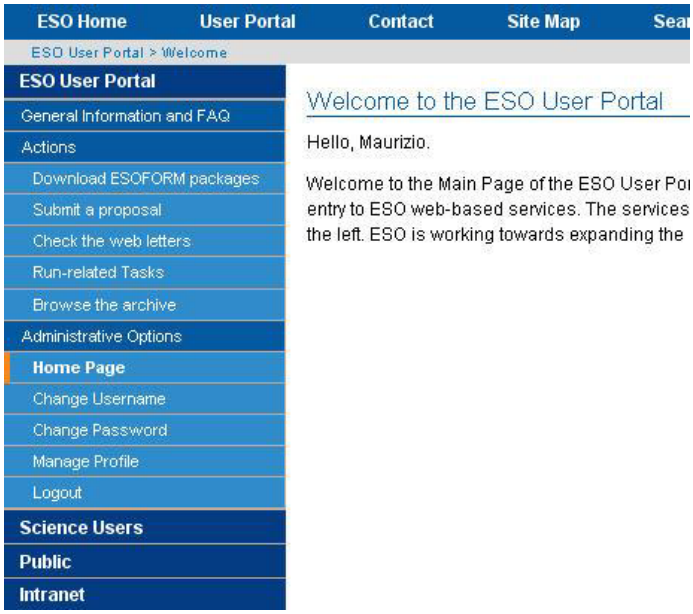

Figure 1: Fragment of the User Portal "welcome" page: the navigation bar on the left-hand side includes all services available to this particular user

[1] amchavan@eso.org
[2] http://www.eso.org/UserPortal

Not only did we unify and harmonize ESO's pre-existing offerings, but we are now offering new features that were made possible by the Portal itself. For instance, Principal Investigators can now have direct access to their proprietary data through the ESO Archive interface.

All in all, the User Portal was quite a successful project, but it's interesting to understand why the development effort took longer than expected and what we could have done to complete it sooner and at a lower cost; it's also important to see what we did well and should do again next time. This paper will try to list some of the lessons we learned, sometimes painfully, during the process of bringing the Portal to the public. Some of the problems we faced originated from historical conditions and organizational issues, although resource and technical problems did have an impact on the schedule as well. Relevant issues include:

- unifying (merging and harmonizing) existing user databases turned out being a major enterprise: international address data is intrinsically messy and difficult to systematize
- providing backwards compatibility for existing applications was more difficult than expected; we eventually resorted to modifying more tools than we had planned
- we chose early on not to adopt any existing Portal technology; a different decision might have saved us time and development effort
- the Organization did not perceive the full scope of the project until well into the project's life
- planning was made difficult by internal operational constraints, lack of committed resources and unified project management

2. PRE-EXISTING USER DATABASES

Two separate science user databases had evolved over time at ESO, and remained in use until the User Portal was introduced.

The so-called "proposal" user database was automatically populated by the ESO Proposal submission system (*PHRS*). While processing proposals, information about the Principal and Co-Investigators (PIs and CoIs) was extracted from the proposal text and stored in the database. The system was taking steps to avoid creating a new database entry if the user was already known (that is, if he or she had already submitted proposals in the past). However, this check was based on the investigators' first and last names and the program was incapable of understanding whether "Maurizio Chavan" and "M. Chavan" represented the same person, or whether the latter really stood for some "Maurice Chavan". As a result, user information was inaccurately stored in the database, mostly resulting in duplicated entries; more rarely, the system got confused and a proposal ended up being assigned to the wrong PI[3]. For that reason, a tedious manual database cleanup phase was needed after every proposal submission phase (every six months), with the goal of restoring the quality of stored user information. Being a manual process, the cleanup phase was not guaranteed to be complete or fully accurate; over the years a number of duplicate or incorrect entries accumulated in the database and they had to be dealt with during the unification process. Before being merged into the Portal user database, the "proposal" database included over 4700 entries.

The so-called "archive" database, on the other hand, included entries for anybody who ever registered with ESO to download files from the Archive (including VLT, La Silla and HST data). It was managed by the registered users themselves, via a Web-based interface (as it came to be with the User Portal), and as a result it included fewer duplications. In its final form the "archive" database included about 4200 entries, many of which represented users who were also present in the "proposal" database.

Not surprisingly, the two user databases had a similar layout, including first and last name, e-mail and postal addresses, and the name of the institution the user was affiliated with – their so-called "affiliation". However, while the affiliation of an Archive user was a single piece of information, in the "proposal" user database a user could be associated with

[3] In one case, two homonymous Swedish astronomers were working at the same institute; they were accidentally "merged", and these changes had to be undone. It later turned out that they shared the same bank, and the bank also merged their details – we were not the only ones with user database problems.

multiple affiliations – in fact, it was an attribute associated to a proposal (for statistical purpose, ESO needed to keep the history of all institutes a PI or CoI was affiliated with while observing with ESO telescopes). Other than that, "proposal" users were identified with a number, while "archive" users had an alphanumeric identification code; we chose to give an alphanumeric ID to the User Portal accounts as well.

Both databases were implemented with the help of a commercial relational DBMS.

2.1 The migration procedure

The new User Portal database had to include entries from both databases, and we identified the following requirements for the migration procedure:

- the history of proposal submissions and data requests should be maintained
- users who had a "proposal" and an "archive" account should be identified and given a single User Portal account (with the "archive" ID)
- all duplicated entries should be merged
- the most recent available information (affiliation and postal address, etc.) should become an account's default

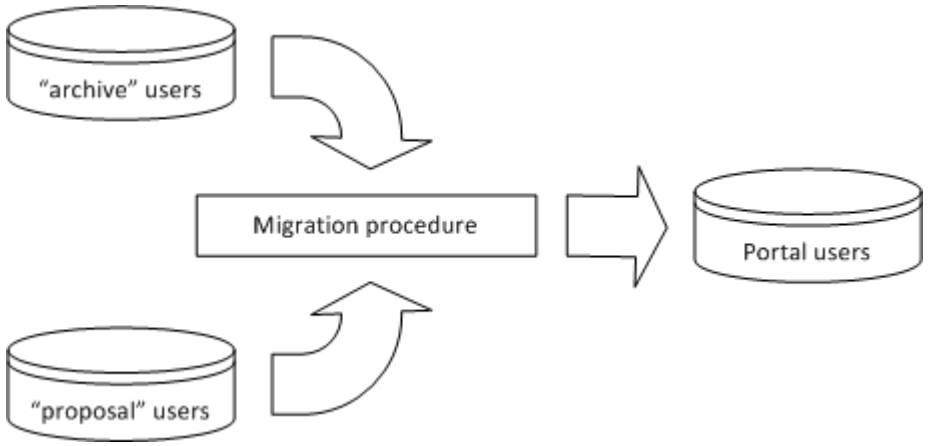

Figure 2: Data-flow of the migration procedure

A common goal of these requirements was to minimize the impact of introducing the User Portal. Early on, we identified user acceptance being a major risk driver, and we strived to make it as easy as possible for our users to move on to the new Portal.

The initial version of the migration procedure was a deceptively simple set of SQL scripts. We used e-mail addresses as "unique identifiers" for users in both databases, obtaining (what appeared as) good results with limited effort. On closer examination, though, it became apparent that the process was only partially successful and the migration requirements were not met. The SQL scripts we had were not robust and flexible enough.

More specifically, an unforeseen large number of "special cases" were found – in the best "80/20" tradition, dealing with a minority of special user cases turned out to take most of the development time. The initially simple SQL scripts grew to unmanageable proportions and had to be re-engineered several times.

Lesson learned: the issue of user data migration should have been analyzed more thoroughly up front.

Lesson learned: in any case, we should have used some other language from the very beginning, as SQL is not very suitable for specifying complex algorithms.

A parallel set of verification scripts had to be developed as well, to make sure that no data was lost in the migration process, and all requirements were met. As with all software, the verification scripts had to be verified themselves to make sure they worked as expected.

An added complication came from the fact that the project development phase stretched over several proposal submission periods (see sec. 6**, "***Planning issues***"**). Semester after semester the "proposal" user database grew, and the number of special cases grew with that; each special case required further refining of the migration scripts. In the meanwhile, new users registered almost daily with the ESO Archive, potentially compounding the problem. In short, the migration procedure was addressing a moving target, and there was no guarantee that today's solution would still work with tomorrow's data.

In the end, we had to give up the goal of a fully automatic migration; it was simply not feasible. We accepted that a few hundred cases had to be dealt with manually, post-migration, before we could open the Portal – and for that problem size, manual checking would be a manageable approach. In fact, even after going live, some 550 as-yet-unidentified duplicate accounts were identified and had to be merged[4].

Lesson learned: we should have given ourselves a less ambitious goal than a fully automatic migration process, and better tools for manual fix-ups.

2.2 Managing affiliations

In both the "proposal" and "archive" user databases the name of a user's affiliation was a free text field. ESO wanted to have better control over affiliations and countries (again, for statistical purposes); for that, we needed to identify a set of "official" research institutes and map each existing institute name to one on the official list. Accounts in the User Portal database were to be associated to one or more of the official institutions.

The roster of all official research institutions was initially compiled by taking into account any institute name found in either user database. The roster was then broken down per country – from Argentina to Venezuela – plus a special list for institutions like Gemini or IRAM which have offices in several countries. Lists were then "fleshed out" by adding all other known astronomical institutes: eventually, we had over 70 lists, ranging in size from one institution (*e.g.* Armenia) to several hundred (*e.g.* the United States of America). For each institution an official name and postal address had to be found, usually browsing their Web pages or consulting the Membership Directory listing of the American Astronomical Society; we often needed the assistance of colleagues knowledgeable with a specific country to disentangle the confusion of departments, offices and divisions of the major institutions.

We then had to map all known variations of an institute name to its official denomination: we had for instance 59 variations of "*Observatoire de Paris, Site de Meudon*", ranging from "*OPM*" to "*LESIA*" to "*Pairs Observatory (Meudon)*" – and yes, that was "Pairs", not "Paris".

The lists were maintained as Excel spreadsheets; version control was obtained by attaching them to the project's Wiki. We needed a script to extract the raw institution information from the user databases into the spreadsheets, and another script to load the spreadsheets back into the User Portal database. Moreover, the institution lists changed over time just like the users themselves[5], so the mapping had to be updated over and over.

Lesson learned: coming up with a complete and valid roster of official institutions implied much more manual work than we originally anticipated.

Lesson learned: quick-and-dirty solutions like spreadsheet-based data entry seldom scale up nicely. We should have eased the problem by providing a database-oriented interactive tool for maintaining the institution list, avoiding the spreadsheets altogether.

---

[4] That effort involved a database dump, followed by *awk* script processing, followed by (in the case of questionable duplicates) interaction with the users – no small effort.

[5] Individual spreadsheets underwent up to ten revisions in the course of the project.

3. BACKWARD COMPATIBILITY

The "proposal" and "archive" user databases were fundamental for several mission-critical ESO applications, including Observing Proposal submission and the interface to the Archive; the User Portal could not inversely affect any of those tools. Most applications accessed user information by means of SQL queries embedded in the code, and while some code changes were deemed necessary since the beginning[6], full backward compatibility for most applications was an important design goal. It was very impractical to find and correctly update all user-related SQL queries in all current applications, including software written in Java, C/C++, Perl, Python, bash and more.

More recent ESO software shares a common database access layer, where all persistency services are centralized. Should the database structure change again, only that layer will need to change; applications will remain unchanged.

Lesson learned: decoupling your persistence code from the application code is a good idea and a valid investment.

The similarities between "proposal" and "archive" user database layouts made us opt for a solution based on SQL "views". Views are virtual relational tables that are built on top of actual tables, and are a standard offering of many RDBMS vendors. The idea was that we would replace existing "proposal" and "archive" user database tables with views based on the User Portal database. Because of the particular technology we used, those views could only be read-only: an analysis of all existing procedures revealed that it wouldn't be a problem. (See Figure 3.)

It later turned out that the analysis had been incomplete, and some applications did need to alter user information. We also found out very late in the development process that one critical application would break if user IDs ended with a number (like "*mchavan126*"); we had to engineer that constraint back into the system.

In the end, we needed to modify more applications than we expected, adding to the development cost of the project and delaying its release. Since some of the issues were discovered relatively late, changes had to be rushed in, adding pressure on developers and testers alike.

Since many applications had to be re-released for the User Portal, we ended up having a rollout phase of the *big-bang* type, one of the main Portal challenges described in [3] (see sec. 6, "Planning issues", for a description of how we looked for candidate release dates). Our rollout required extremely carefully planning and tight control; in the end it went very smoothly, in spite of some last-minute unexpected glitches.

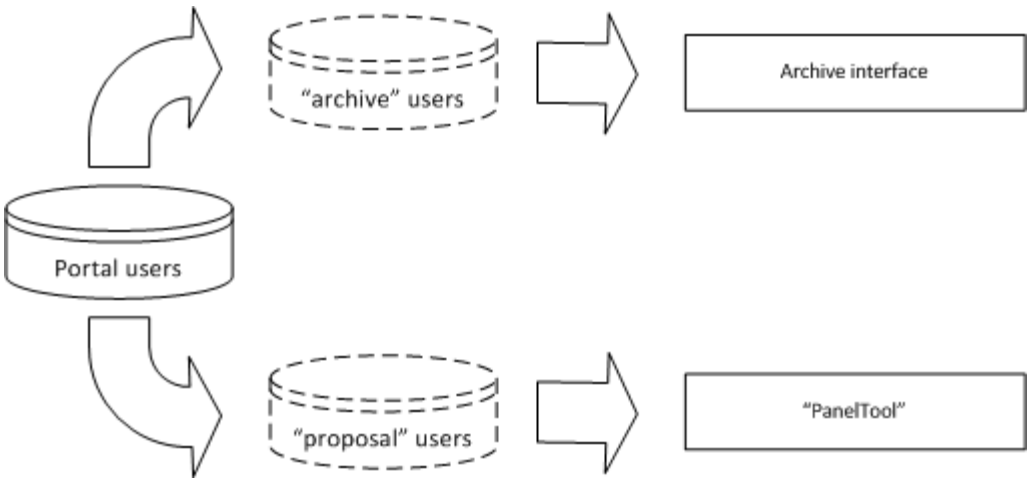

Figure 3: The database views simulating the pre-User Portal user databases were drawn with dashed lines; some example applications reading user data are shown as well. In practice, the Views invert the logic of the migration procedure

[6] For instance, the Proposal submission system was to be made available only to registered users, and should not update the user database anymore; the Phase II system was to accept alphanumeric user IDs, etc.

Lesson learned: thorough end-to-end test should be performed very early on, and they should be repeated as part of every software release. (We identified that need early on in our risk analysis, but were not consequent enough.) No matter how careful your initial analysis is, you will most likely find unexpected quirks in the system. End-to-end tests should be performed on a software and hadware configuration that's as close as possible to the production system[7].

Lesson learned: build abundant slack in your schedule! You really never know what's going to happen when you try integrating very diverse tools under a common umbrella.

Lesson learned: tight planning of critical phases pays out nicely.

4. TECHNOLOGY CHOICES

When we set out designing ESO's User Portal we deliberately chose not to use any available Portal technologies. The decision was based on a number of reasons, including the fact that we had a predefined design for our user databases and needed to "portalize" applications written in a number of different languages, over which the development team had no control. No existing Portal offering seemed to be able to integrate a variety of technologies and pre-existing databases; in any case, we wanted to remain independent of any particular product.

The main reason for a home-grown Portal approach was however that ESO was at that time in the process of selecting a Contents Management System for the Organization's Web. It was felt that proper CMS selection was paramount, and we wanted to impose as little constraints as possible on the system to be chosen – for which very many critical requirements had already been expressed. We only asked that the CMS, which was also to provide a comprehensive presentation framework, would be able to accommodate Java applications.

As we started out rolling our own Portal infrastructure we tried to keep it small and simple, knowing that one day it would be replaced by something else. However, as the development process dragged on and more features were required (see sec. 6**, "***Planning issues***"**), the infrastructure itself had to be extended and ended up absorbing a substantial fraction of our development resources. In addition to that, implementing even a bare-bones Portal infrastructure meant that we had to cope ourselves with rather complex technical problems, like *Single Sign-On* (and -off).

In retrospect, it may well be that we could have saved ourselves time and effort[8] by basing our development on some available Portal product instead of writing our own infrastructure.

Lesson learned: take time to shop around before you decide to do your own development, and remember that while most projects start small and simple, many end up growing beyond expectations.

5. ORGANIZATIONAL ISSUES

When the User Portal eventually went live it was appreciated by just about everyone, including ESO's user community and up to the highest management levels, proving that it had been a good idea whose time had come. The benefits it provides to end users justify the investment that was necessary, and more features will become available in future Portal releases that would have been impossible to offer without the the work we did. The Portal is now taken for granted by users and management alike, which is arguably the best recognition it could get.
Possibly, development would have been smoother if that recognition had been there earlier.

ESO was present on the Web for many years without a User Portal; arguably, the Organization could have continued for some more years without a unified access point for its science services. The project was started based on the vision of a single manager, and it continued as an initiative of the Data Management Division[9] after that manager left. For a major part of its life, the User Portal was perceived as a software project: its scope, how far it reached into the Organization, was not immediately apparent to most people outside the development group.

---

[7] Only as we went live did we find out that for one critical feature the target production system was considerably slower than the development and integration systems.
[8] Of course, it's hard to estimate now how much we would have saved.
[9] It is now called *Software Development Division*.

This lack of visibility had a number of consequences.

- The project was often given second rank with respect to minor but established software development and maintenance tasks: when some development bottleneck happened, development and testing resources were diverted from the User Portal to other operationally critical software tools. That is not just a matter of development time: if developers switch between different projects they find it harder to keep concentrated on the task they're trying to complete, and become less productive.
- An organization's Portal can by no means be the effort of a single division, but it took several months before all needed players could be taken around the same table, and more time yet before the scope and implications of the project became clear to everyone. The first groups to be involved were development teams and other technical people; operations and management followed later.
- That made the decision making process very slow. Governance issues could not be decided until very late; for instance, the "proposal" and "archive" user databases were maintained by separate operational teams, but who should then maintain the User Portal database? Coordination issues, especially those related to the critical rollout phase, were left pending until dangerously late. Ten months into the project we still didn't know whether we should keep the official list of institutions (see sec, 2.2, "*Managing affiliations*"), nor whether a proposal's co-investigators had to be registered users or not – a major issue for the proposal submission system in the User Portal era.
- No unified project management structure could be established, until close to the rollout phase; there was no unified quality assurance plan, either. Coordination was done on a best effort basis – luckily individual developers and testers developed a strong sense of attachment to the project, which compensated for those shortcomings.

Lesson learned: win management support early on, up to the organization's top management. Don't launch the project until you're sure all institutional stakeholders are aware of the risks, costs and benefits involved. Make sure all involved parties talk to each other throughout the project's life: nothing should be taken for implicitly granted.

Lesson learned: make sure the project is properly managed, with support from all involved line managers.

On the positive side, the collaboration of technical and operations staff worked rather well. Regular team meetings (including all involved parties) were established a year into the project, and proved fundamental in promoting a common understanding of the project's goals and course[10]. A byproduct of improved communication is that a common language was established early on, with operations people quickly taking up engineering jargon and engineers understanding operational constraints.

Lesson learned: keep the project team in close contact, promote exchange between all involved parties.

6. PLANNING ISSUES

As indicated above, the first version of the conceptual document describing ESO's User Portal was circulated at the end of 2004. That document went through a number of iterations, slowly reaching a larger audience, and after about six months the software team started developing "use cases" for the project and an initial design of the database. After an initial shot at releasing the system in April 2006, the first "realistic" release date was set for August of that same year; that release date could not be met.

In hindsight, it is clear to see that a consequence of the project's uncertain standing – with respect to other development efforts – was that developers ended up working only part-time, being continuously distracted by other more pressing tasks.

Lesson learned: make sure you have the proper resources and they are committed to the project. Don't allow "urgent" tasks to override "important" tasks.

[10] We later had to establish a separate set of smaller, more technical meetings, as the project meetings had become too large.

Setting release dates for the User Portal was not easy. No matter how well we planned, the rollout phase would cause a certain amount of disruption of the operational flow, which was unacceptable during certain phases of ESO's semester-based cycle. No-go times included the Proposal and Phase II submission weeks, the weeks around the Observing Program Committee's meeting[11], the annual review, the beginning of the observing period in Chile, etc. As a result, missing a release date meant that the next available slot would be quite some time away, in one of the few available "slow" periods.

The next release date was set to be in December 2006. Those few extra months to complete development were meant to allow the team to complete implementation of the missing features and remove all known bugs from the code, but then another (well known) phenomenon appeared. With a little more time on their hands, developers started to think they could fit some extra work into the schedule, like some important refactoring or a more flexible way to print address labels. On the other hand, scientists thought *they* also had a chance to promote some new feature into the plan, usually as a result of requirements clarification or re-prioritization. As a result, the project plan became a moving target: the extra available time was not dedicated to the original consolidation objectives; instead, we were trying to deliver what effectively was a different system.

Lesson learned: do not allow "*feature creep"*, even if it means rejecting some very reasonable requests. Assure your stakeholders that all important features will be implemented in due time, and don't let them change your to-do list.

We could not make the December 2006 release. At the end of 2006 one of the main User Portal developers left the team, leaving an important void, which took months to fill.

The following release, planned for February 2007, was canceled by management because it would conflict with some operationally critical phase; we missed a release planned for May as well. Along the way to the first release we de-scoped some of the original requirements (early on, an LDAP interface to the user information; later, support for "Delegation of Responsibilities"[12] and "Science Teams"; finally, the "Address List" feature, etc.).

We also found out quite late that we could not meet the release dates, which didn't help improve the credibility of our estimates.

Lesson learned: find a way to measure progress, and keep that measure up to date. A four month gap between releases is too long to keep the pulse of the project.

A very happy team finally saw the User Portal go live for the entire ESO community on November 15th, 2007.

7. THE FUTURE

Development of the User Portal continues. We're concentrating in making available some important features that had to be de-scoped from the original plan, plus other potentially interesting developments.

- The "Address list" will allow users to define multiple delivery addresses for their proprietary science data, and have it *e.g.* delivered to a collaborator
- "Delegation of Responsibilities" will allow *e.g.* a Principal Investigator to delegate the preparation of Phase II material to some collaborator, while retaining full control over the program and its resulting data
- "Science Teams" will be created and maintained under the responsibility of a team leader. It will be possible for teams to *e.g.* submit Observing Proposals
- The User Portal will be integrated in ESO's upcoming Contents Management System

[11] That is ESO's name for its time allocation committee.
[12] This refers to the case where a Principal Investigator may want to delegate some operational responsibility to a collaborator; for instance, the preparation and submission of an Observing Time Proposal.

8. CONCLUSIONS

The User Portal is a project we're proud of, and we certainly learned some important lessons during the course of its development. Almost all of them seem very obvious from today's vantage point, and they might have been obvious during the tight years of 2006/07 as well, had we allowed ourselves to take a break and look at what was happening[13]. Instead, we worked very hard trying to reach the goals we had set ourselves, and lost track of *how* we were trying to get there. So, perhaps this is the most important…

…*lesson learned*: keep some distance from your project. Do not allow the day to day pressure to distract you from keeping your process under constant scrutiny.

ACKNOWLEGEMENTS

The authors would like to thank Dave Silva, for his original vision; Tim Canavan, for his invaluable contributions to the project; and Thomas Bierwirth, for reviewing this paper's draft and providing insightful feedback.

REFERENCES

[1] A. M. Chavan, L. E. Tacconi-Garman, M. Peron, F. Sogni, T. Canavan, P. Nass "Unifying access to services: ESO's User Portal", in Observatory Operations: Strategies, Processes and Systems, edited by David R. Silva and Rodger E. Doxsey, Proceedings of SPIE Vol 6270, 6270OI (2006)

[2] Chavan, A. M., Peron, M. Anwunah, J. Canavan, T. Dorigo, D., Kornweibel, N. and Sogni, F, "Moor: Web access to end-to-end data flow information at ESO", in Optimizing Scientific Return for Astronomy through Information Technologies, Peter J. Quinn and A. Bridger, eds., Proceedings of SPIE Vol. 5493, pp. 54-62 (2004)

[3] N. Sharma, "Key Challenges in Portal Adoption", on The Server Side; http://www.theserverside.com/tt/articles/article.tss?l=KeyChallengesinPortalAdoption (11/2007)

[13] In fact, some of those issues were very visible from the very beginning, but there was little one could do about them,