Software Sustainability Institute

ICAT Collaboration Sustainability Assessment

Project Title	ICAT
Document Title	ICAT Collaboration Sustainability Assessment
Authorship	Steve Crouch [Author]; Alistair Mills [Editor]
Document Filename	ICATCollaborationAssessment-v1 0.docx
Document Version	1.0
Document Number	ICAT-1
Distribution Classification	SSI-internal/ICAT-internal
Distribution List	*SSI Project Team; ICAT Project Team*
Approval List	*Project Team*

This report is based on interviews with ICAT stakeholders, which took place during January 2012. The University of Edinburgh on behalf of the Software Sustainability Institute does not accept liability for acting on any of the recommendations contained in this report.

University of Southampton

Contents

Table of tables

Table of figures

Acknowledgements

Firstly, I thank all of the interviewees for donating their valuable time to this assessment. Thanks also go to Alistair Mills from the ICAT team for efficient and effective preparation and organisation of the interviews and the overall framework for assessment. Thanks should be given to Harjinder Chahal from the ICAT team for his welcome, prompt and willing assistance throughout the course of the assessment.

Executive Summary

Both the ICAT collaboration and the ICAT software are at a critical point in the ICAT life cycle. To date, the mechanisms in place for the management of ICAT have met the needs of the partners in the collaboration. However, ICAT is facing the problems associated with the successful and increasing uptake of the software. There is general agreement, and specific requests from stakeholders, that the mechanisms expand to support a larger community of users. The driving force for change comes from the PaNdataODI and the EuDAT projects which are evaluating ICAT as a metadata cataloguing solution for widespread use in Europe.

The principal source of data for this assessment was a set of fifteen interviews. These interviews led to the identification of a set of observations, leading to a set of recommendations for action by the collaboration team in general, and the eScience ICAT team at STFC in particular.

The key factor for the future success of ICAT is that the collaboration provides assurance to the stakeholders that ICAT will continue to meet their needs.

There are some major positive observations about the current state of the collaboration. One is that the core ICAT team is already well aware of many of the issues identified in the interviews; the team is asking the right questions on sustainability at the right time during the collaboration's evolution, and that the team accepts that changes in a number of areas are required. In addition, the issues identified are generally those that one would expect from a project at ICAT's level of maturity; the issues include the successful management of an increasing number of requirements and stakeholders, ensuring a higher level of involvement and promotion with a growing stakeholder community, and ensuring that the development processes, policies and infrastructure can meet the needs of the stakeholders.

There are several important observations. Firstly, ICAT is an integral part of the operational processes of the stakeholders, who have a dependency on the software [CO1]. The collaborative processes and mechanisms in place have met the needs of the collaboration well [CO2]. However, it is clearly important that these mechanisms expand to ensure a sustainable future for the collaboration [CO3, CO4]. In addition, there is a need for the definition of more governance, process and policy, to manage the expected increase in use [CO5, GO2, GO3, GO5 and GO6].

The assessment provides a set of recommendations of which the following are the most important: define the governance of the collaboration [R1]; create a well-defined induction process for new collaborators [R13]; improve and streamline the requirements ingest process [R16]; expand, improve and promote the road-mapping activity [R17]; develop policies for deprecation, code contribution and software release [R22, R23, and R24]; allocate effort to assurance and promotion [R9, R11]; create a group of external product testers [R18]; engage actively in facility roll-out plans [R19]; create a capability to provide mentors for new collaborators [R7].

1 Introduction

1.1 Overview

The aim of the assessment was to determine the sustainability of the collaboration that continues to develop around the ICAT software, which is actively developed and maintained by the e-Science group within the Science & Technology Facilities Council at Rutherford Appleton Laboratories.

The assessment was conducted by Steve Crouch, a senior consultant from the Software Sustainability Institute (SSI), and the preparation and organisation of the assessment was conducted by Alistair Mills, project manager at STFC. The assessment included interviews with 1 principal investigator, 1 project manager and 5 developers from the ICAT team within the e-Science group at STFC, 6 users across ISIS, DIAMOND, SNS and ELLETRA, 1 user/developer from CLF and 1 developer from DIAMOND.

1.2 Assessment Objectives and Scope

The assessment was conducted between 16th and 20th January 2012, and had the following objectives:

- To have SSI make an assessment of the sustainability of the ICAT collaboration;
- To produce a list of actions which the ICAT collaboration can implement to operate more effectively;
- To produce a paper or report which can be discussed at the ICAT meeting in Grenoble in March 2012.

Based on this assessment, the ultimate objective is to improve the efficacy of the ICAT collaboration.

1.3 Assessment Methodology

The assessment involved a week of on-site interviews at Rutherford Appleton Laboratories with the various stakeholders involved in the ICAT collaboration. These stakeholders included representatives from the ICAT development team as well as on-site and external users of the software from a variety of facilities. These interviews were arranged in advance by the on-site assessment organisers, and were accompanied by software and infrastructure demonstrations by the ICAT team. These activities were followed by a period of off-site analysis of the interview material and the writing of this report. The estimated effort and perceived importance figures are subjective estimates made by the assessor.

1.4 Involved Stakeholders

Below lists an overview of the ICAT collaboration stakeholders involved in the assessment process:

Table 1: Participants by Stakeholders

Stakeholder	Participants	Roles
e-Science, STFC (Science & Technology Facilities Council, RAL, UK)	7 (internal)	1 principal investigator, 1 project manager, 5 developers
ISIS (RAL, UK)	2 (local)	2 users
DIAMOND (RAL, UK)	3 (local)	2 users, 1 developer
CLF (Central Laser Facility, RAL, UK)	1 (local)	1 user/developer
SNS (Spallation Neutron Source, US)	1 (external)	1 user
ELLETRA (Italy)	1 (external)	1 user

References to the stakeholders are by acronym/abbreviation, as above. *Internal* refers to those participants that are part of the core ICAT team, *local* refers to those on-site at RAL, and *external* refers to those off-site. Comments from the assessor are tagged as 'SSI', where no corresponding comment exists from one of the stakeholders.

2 Observations

2.1 Collaboration Overview

2.1.1 Overall

CO1. ICAT has become an integral, even critical, part of the operational processes of the stakeholders, raising a clear dependency on the software for the foreseeable future.

CO2. The processes and other mechanisms currently in place for managing the ICAT collaboration have, to date, been successful in most cases in addressing the needs of the stakeholders and the overall collaboration. This has lead to the large majority of stakeholders being very satisfied with the software, management of requirements and level of support provided by the e-Science ICAT team.

CO3. Many of the external stakeholders have identified the importance of having, and promoting, a sustainable future for the collaboration and the software. This would offer assurance that their requirements will be met in the future.

CO4. There is a concern, identified by both external stakeholders and the e-Science ICAT team, that with the clear possibility that PaNdataODI will recommend the adoption of ICAT, a more rigorous approach is necessary to managing requirements and potential requirement conflicts.

CO5. There is clear acceptance within the e-Science ICAT team that more governance, process and policy will be required to manage the broader uptake of ICAT that is expected in the future.

2.1.2 Funding

CO6. Funding for the ICAT team and external stakeholders does not appear to be an issue for the foreseeable future, although a high turnover of staff within the ICAT team has occurred previously [e-Science, CLF].

CO7. One issue highlighted is that the ICAT collaboration should not be too dependent on PaNdataODI to survive. STFC must commit effort to leading this in the longer term to continue the work. There is a requirement to make sure ICAT will not be dropped; this level of assurance is important [ISIS].

CO8. One suggestion is to have a subset of the ICAT team actively looking for future external funding [ELLETRA].

2.2 Governance

2.2.1 General/Leadership

GO1. At least one ICAT user is very satisfied with the effort from the e-Science team [ISIS].

GO2. There has been a light touch to governance of the ICAT project to date, realised by the ICAT team and external stakeholders, which has been largely sufficient [DIAMOND, ISIS, e-Science].

GO3. However, there is a realisation that more governance is needed with other projects coming into the collaboration e.g. ELLETRA, PaNdataODI in general [ISIS, e-Science].

GO4. The collaboration is currently undergoing a big transition, since e-Science has become involved again after a hiatus. The ICAT team is taking more control of the leadership of the collaboration – and this needs to continue [ISIS].

GO5. It would be valuable to pin down the governance of the collaboration with a defined governance policy. This has been tried before, but it was difficult to get positive engagement [ISIS].

GO6. STFC needs to commit effort to leading the collaboration, to ensure it will not be dropped; ICAT users need this assurance [ISIS].

GO7. e-Science should take on the organisation of the Face-to-Face meetings [ISIS].

GO8. There have been some difficulties with participation in meetings; some of the collaborators have difficulties with the time of the meetings [e-Science].

GO9. There is an understandable general concern to look after the requirements of a new user organisation, but this needs to be balanced with existing stakeholders' concerns. There is a need to take care in managing the balance between the roles of leading the ICAT collaboration and managing a deployment project [ISIS].

2.2.2 Structure and Stakeholder Involvement

GO10. It is apparent that the ICAT team is very cohesive and there is a high degree of communication within the group on organisational and technical issues [SSI].

GO11. The fortnightly collaboration meeting has been widely praised as a highly inclusive, polite, welcoming environment, a positive experience with excellent interactions and high responsiveness which is generally open in communications [SNS, ELLETRA, ISIS, e-Science].

GO12. The current structure is generally OK, but is fairly untested in international collaboration; it is very RAL-focused at the moment. Internationally, this could cause problems especially with time-zones [e-Science].

GO13. At the moment, there is one meeting to discuss both technical aspects and policy; this has been fine up until now, but which should ideally be split [ISIS]. With more potential partners, there is a need for a higher-level body, such as a Steering Committee for the collaboration [ISIS, e-Science]. Such a body would need representation from PaNdataODI, but it would need to be decided how this would work with other collaborators as well e.g. SNS [ISIS].

GO14. At the moment, the collaborators in DIAMOND see themselves as third party users, albeit with requirements that are being taken into account – a basic supplier/customer relationship. However, this probably needs to change as requirements increase and evolve [DIAMOND].

2.2.3 Joining the Collaboration

GO15. Joining the collaboration as a new collaborator is generally a smooth, open process [DIAMOND, e-Science, ISIS].

GO16. There is a question about how a new partner organisation should join the collaboration. This seems generally good at the moment, but in the future it may be necessary to clarify which roles are expected, so that the partner organisation can assign the appropriate individuals [e-Science].

GO17. How should a new collaborator be involved in terms of their users and developers? [e-Science]

2.3 Community and Outreach

2.3.1 ICAT Group/Stakeholder Communication

CM1. Communication to date has been largely effective, and continues to improve, with the ICAT team well balanced to listen and respond to stakeholder requirements and concerns [ISIS, e-Science, SNS].

CM2. Specifically, Tom Griffin from ISIS has adopted a critical and successful role in many areas, including ISIS representative, initial external user contact (e.g. SNS), product champion, technical support (e.g. SNS, DIAMOND, CLF), assistance in maintenance of the GoogleCode ICAT project, and external ICAT tester (e.g. for e-Science). This has been

necessary in the past with e-Science's hiatus in ICAT development. However, Tom's role will change with the ICAT team adopting a greater leadership and organisational role, and as he moves towards a more customer-oriented role within the collaboration [ISIS, DIAMOND, e-Science].

CM3. This observation has been removed from the report.

CM4. Less formal Skype chats once or twice a week would be useful, arranged and focused at accomplishing specific tasks with the right people on the call; these calls would keep others up to date on more specific issues [SNS].

2.3.2 Feedback from Users

CM5. The ICAT core team would like to have more feedback from users on installation, use, general issues, or anything else e.g. what do people think about ICAT 4? So far, they have only heard back from SNS [e-Science].

2.3.3 User Outreach

CM6. There is an issue about the lack of general awareness of ICAT outside of the collaboration e.g. initially, PaNdataODI did not know about ICAT [e-Science].

CM7. Who is currently using ICAT outside of those already known? [ISIS].

CM8. Should the ICAT team be consulting people who do not currently use ICAT to determine their requirements? [ISIS]

2.4 Process

2.4.1 General

PR1. The current development process is reasonable, with its low volume of requirements at the moment – easy to manage [e-Science].

PR2. There is a possibility that with a greater number of collaborators and requirements, a conflict resolution process may be required [e-Science].

PR3. There is an awareness that as new features are requested in the future (from a wider community) this will lead to increasing complexity in the ICAT software and potential scalability problems with its development [e-Science, ISIS].

PR4. A concern exists that the development of the software is currently too 'open-ended' [CLF].

PR5. The collaboration is beginning to tighten up the processes to avoid future mistakes that have occurred with external users in the past [e-Science].

PR6. There is agreement that more process and policy will be required, while focusing on a 'light touch' development process [e-Science, ISIS].

PR7. Generally, processes are internal and neither externally visible nor understood [SSI].

2.4.2 Requirements Process

PR8. There is general agreement that the methods of handling requirements to date have been sufficient with the current group of stakeholders [e-Science].

PR9. The method of managing incoming requirements is largely an ad-hoc process e.g. via email, GoogleCode issue tracker, meetings [e-Science].

PR10. However, there is an acknowledgement that the software is diversifying as more feature requests come in, with some stakeholders not explicitly wanting the new features [e-Science].

PR11. Correspondingly, there is also a clear divergence of user communities, which is an issue [CLF].

PR12. Software changes made so far have been for the better, but will need to be considered much more carefully in the future [ISIS].

PR13. There is concern about how ICAT 4 development will be influenced by big detectors and subsequent impact on how ICAT operates within CLF [CLF].

PR14. It is likely that PaNdataODI will adopt ICAT, and requirements from the partner institutions will need to be managed [ISIS, ELLETRA]. The analysis of these requirements to address conflicts represents a challenge, with different facilities having different requirements as well as different core business. The mutual exchange of requirements from other facilities is very important [SNS].

PR15. Discussion and compromise about requirements with stakeholders has happened to date, but future stakeholders need to be involved in this process [e-Science]. There needs to be assurance that the changes to meet new requirements will not break the support of existing deployments [ISIS].

PR16. Some previous changes to ICAT have been detrimental in the last 6-8 months (e.g. tape support for DIAMOND, incompatible API changes) [ISIS].

PR17. This observation has been removed from the report .

PR18. Requested changes are implemented and are well represented within the repository. However, it is unclear how requirements are prioritised within the ICAT team [SNS]. Prioritisation could be improved; it is currently based largely on what can be achieved [ISIS].

PR19. There is recognition within DIAMOND that new requirements are increasing, and there is a need to pass them on to the ICAT team and to ensure that they are properly understood [DIAMOND].

2.4.3 Future Planning

PR20. A process exists for discussing requirements with collaborators, followed by internal prioritisation. Engagement for this occurs within the fortnightly collaboration meeting as well as an ICAT roadmap meeting once per year [e-Science, ISIS].

PR21. However, some groups seem to be more involved in this process than others [SSI]. Some consider themselves well involved at present [some within DIAMOND, e-Science, ISIS, SNS]; others are either unaware that there is a process, or indeed a roadmap, or are not appropriately involved [DIAMOND, ELLETRA, CLF].

PR22. Assuming ICAT is adopted, there is a list of requirements expected from PaNdataODI's D4.1 survey deliverable that will need to be managed by ICAT [ELLETRA].

PR23. There is very strong support for a more public, advertised, transparent and community-based road-mapping process involving the existing and future stakeholders [DIAMOND, ISIS, ELLETRA, SNS].

PR24. Future developments in terms of a roadmap (and at a lower level software release changes), need to be widely advertised in a timely manner; particularly to avoid any potentially harmful changes [DIAMOND, CLF].

2.4.4 Test/Roll-out Process

PR25. The test and roll-out processes include quite a good and repeatable informal test process, with formalisation of these test procedures beginning [e-Science].

PR26. ICAT testing external to e-Science is also achieved (e.g. through Tom Griffin for ICAT 4.0) [e-Science, ISIS]. External testing for software is regarded as a very valuable activity [SSI].

PR27. Upgrading ICAT is necessarily a careful process for DIAMOND, given very limited and strictly scheduled downtimes. Previously, this has been accomplished by parallel phasing in/out of ICAT versions 3.1/3.3, but with the greater data volumes now experienced keeping each version lock-stepped would be a challenge [DIAMOND].

2.4.5 Release Process

PR28. Progress has been made towards standardising a release process with an approximate six-monthly release cycle, which has been beneficial in avoiding code repository check-in issues that have occurred previously [e-Science].

PR29. There is a need to be careful of divergence of the code base, and to avoid forking of implementations that are tailored to individual projects [DIAMOND].

PR30. Some stakeholders rely on an informal, direct release process, with specific releases addressing their requirements emailed to them [ELLETRA, DIAMOND]. This has functioned very satisfactorily in the past, but will become unsustainable as ICAT's community increases [SSI].

2.5 Policies

2.5.1 Deprecation

PO1. Deprecation of features needs to be handled very, very carefully (e.g. changes to the ICAT API between v3 and v4). ISIS will inevitably need to phase in/phase out parallel deployments of different versions. This should not be necessary if deprecation is handled properly [ISIS].

PO2. There is agreement that more process and policy will be required e.g. deprecation policy is currently missing [e-Science].

PO3. How features are deprecated is very important, with a clear need for a deprecation policy [ISIS, SNS, DIAMOND].

2.5.2 Code Contribution

PO4. The moderation of code being checked in to the code repository could become an issue in the future [e-Science].

PO5. A code contribution policy would be definitely useful when requirements from PaNdataODI become clear (assuming adoption of ICAT by PaNdataODI) [ELLETRA].

PO6. No policy is in place for code contribution, but it would help to have one [DIAMOND].

PO7. If there were more contributors, there would have to be more control over what they are doing, and the policies would need to be policed. [e-Science].

PO8. As code contributions are at a low level at present, the code is simply sent direct to the ICAT team by email. In the future, the developer at SNS will be happy to check code in referencing created tickets [SNS].

PO9. The assignment of ticket numbers to SVN commits is improving [e-Science, ISIS].

2.5.3 Release Policy

PO10. Notification of release schedules is important for infrastructure upgrades and development [SNS, DIAMOND]. If it did exist, it would definitely be used – critical [DIAMOND]. Advertising release schedules could perhaps be twice a year [SNS].

PO11. When the trust model is due to change for a deployed server, this should be advertised well in advance, with an appropriately timed release of the updated trust-store for clients. This has impacted demonstrations in the past, although issues were quickly resolved [DIAMOND].

PO12. Currently no version change log or release notes are available for ICAT 4.0 [e-Science].

2.6 Product Support

2.6.1 General

PS1. Generally support requests have been handled very well, with quick responses [ELLETRA, DIAMOND, e-Science].

PS2. Some requests have not been handled well [CLF].

PS3. One issue is the level of Java expertise required to understand some of the support replies. The user's expertise should be taken into account [CLF].

PS4. This observation has been removed from the report.

2.7 Infrastructure

2.7.1 Google Code

IN1. There is solid acceptance that GoogleCode is the right infrastructure for the project from those currently using it [e-Science, ISIS].

IN2. It is important that the activity in the GoogleCode infrastructure done not stagnate, as the perception of ICAT as an active project is important for its adoption [e-Science].

2.7.2 Mailing Lists

IN3. Tools are in place, but the collaboration should be making more use of them. There is perhaps too much direct communication between individuals which could be opened up using e.g. Google groups, mailing lists [e-Science].

IN4. DIAMOND collaborations have registered on the mailing lists, but they are a bit quiet; more people should use them, so that there is more discussion [DIAMOND].

IN5. It would be good to have mailing lists to share and solve common problems. We support lots of users with very similar questions [SNS].

IN6. A FAQ and troubleshooting guide would be great for resolving common problems [SNS].

IN7. A good improvement would be a stronger, standard mailing list with experts present on the list to guide users through installation and configuration. This would be useful for those trying it out. It also helps to avoid the issue of forking, since this encourages them to communicate directly with the ICAT team to resolve issues and not modify the code in-house just for them [ELLETRA]. This also would potentially let you know more about your user base [SSI].

2.7.3 Virtual Meeting Mechanisms

IN8. Less formal Skype chats once/twice a week would be useful, arranged and focused to accomplish specific tasks with the right people on the call, and would keep others up to date on more specific issues [SNS].

IN9. Communication via video would be helpful for collaboration [e-Science].

2.7.4 State of Infrastructure

IN10. Many improvements aimed at simplification were introduced in ICAT 4.0. These need to be reflected in a cleaned-up code repository by removing older, unused files [e-Science].

IN11. It is sometimes difficult to find information on the infrastructure [DIAMOND].

2.7.5 Community Support Infrastructure

IN12. This observation has been removed from the report.

IN13. In the existing pre-ICAT system, there was an issue of responsibility for 'the Linux box' in R7 which runs the pipeline software, which previously had not been backed up. An incredibly unlucky circumstance was the failure of two key RAID disks where all data

and manufacturer driver software was lost. Much was subsequently recovered, although this was a hugely expensive recovery. Back-up procedures are now in place. However, this has made the CLF extremely nervous about future support/maintenance of the new ICAT-based system [CLF].

3 Software

3.1 Role

SO1. Software needs a clearly defined role [DIAMOND, e-Science].

SO2. Is ICAT supposed to act as an archive, and a publication database with clearly defined data? There needs to be a clear expectation of what it is supposed to be with respect to other projects. There is a risk of defining it as a silver bullet, where expectations cannot be met and the software is abandoned. For DIAMOND, it is a publishing database, with links to the data archive [DIAMOND].

3.2 General Views

SO3. It is an evolved, stable, usable product that has seen many improvements in recent years [e-Science].

SO4. ICAT is pretty good. During earlier phases of development, care was taken to use as many best practices as possible [DIAMOND].

SO5. ICAT is flexible, and works well for ISIS. However it needs some improvement. In addition, the Swedish road transport facility uses it for managing road transport data, with two MSc students working on it [ISIS].

SO6. ISIS is not generally aware of any problems at all. ICAT seems very stable, appears robust and well written. It delivers on its promise; it is a nice way to get at data for users, and its security is good (previously ICAT did not have this) [ISIS].

SO7. There is good work currently being done on simplification [ISIS].

3.3 Documentation

SO8. There are strong advocates within the ICAT team for good documentation, which is getting better [e-Science].

SO9. Documentation is generally ok/average [DIAMOND, e-Science].

SO10. Some documentation is very CLF-centric [ISIS].

SO11. The best thing they [the ICAT team] have done is the client examples project in SVN. For developers, these examples are richer than any documentation [DIAMOND].

SO12. There is documentation but it could be better; a better introduction to the software would be good. The purpose of the PaNdataODI D4.1 survey is to try to hint at why these data catalogues are important. There is an assumption that users know what they need, but they are not aware of the advanced features of cataloguing that ICAT provides. ICAT could advertise the benefits of the software clearly to raise this awareness [ELLETRA].

SO13. The documentation was originally pretty scant. It is better now, but it is still a bit arcane about how to install ICAT. If you follow the installation documentation, it works [DIAMOND].

SO14. The documentation for the APIs is lacking [DIAMOND].

3.4 Testing

SO15. There is a set of unit tests, which are reasonably good. Continuous level testing (via Jenkins) is present for unit tests on SVN changes [e-Science].

SO16. There is a need for better test suites [e-Science].

SO17. It is not clear how testing is performed at the moment. Continuous testing is important; it is not clear how this operates within the ICAT team. Regression tests are very useful; we should have these in the software as a test suite [SNS].

SO18. There is a need for improvement in error handling; when uploading collections of files using xml-ingest; an error in one file causes all subsequent files to be flagged as in error also [DIAMOND].

3.5 Performance and Robustness

SO19. There are no performance considerations; if it works for DIAMOND, it works for ISIS [ISIS].

SO20. Ingest is slow for putting data into ICAT via web services; this is an issue if there is a lot of data. This problem could become greater in the future. The archiving of previous data is done manually, but live data are currently ingested automatically, unless the live data set is too large, in which case it is done manually. The ICAT team is well aware of this [DIAMOND].

SO21. Despite this, ICAT performance is holding up really well using an Oracle DB cluster at the enterprise level. DIAMOND has well defined use cases, in a well understood environment [DIAMOND].

3.6 Deploying/Developing with the Software

SO22. ICAT 4 is much better than ICAT 2 [SNS].

SO23. With the present installation procedures, we can probably get 80% done, but we have needed the ICAT team for remainder [SNS].

SO24. ICAT is flexible, and it is easy to add new features [SNS].

SO25. The software is manageable, and has been successfully integrated into our own portal [ELLETRA].

SO26. When developing software DIAMOND is generally happy with ICAT as a solution. DIAMOND uses the available WSDL from ICAT web services to build a client JAR file which is enough [DIAMOND].

SO27. The addition of exceptions can currently break code. Some things are not as regular as they should be in ICAT (e.g. the way keys are handled in the database) [e-Science].

SO28. Backwards compatibility and extensibility is important [e-Science, CLF].

SO29. ISIS successfully implemented an interface change to upload data. It is extensible [ISIS].

SO30. Once you understand it, it is straightforward to develop software using WSDL-based clients [DIAMOND].

3.7 Opinions on Future Uptake by Others

SO31. There is quite a high barrier to start using the software [e-Science, SNS, DIAMOND].

SO32. There are lots of dependencies. ICAT is most effective when tightly coupled with business systems; there is a need for better advice and tools to do this [e-Science].

SO33. It is generally good, not a problem, very open [DIAMOND, ELLETRA, ISIS].

SO34. For new external users, they do not perceive any barriers; the mailing list members are friendly and helpful [DIAMOND].

SO35. The software is well placed to meet future demands [DIAMOND, e-Science, ISIS].

SO36. There is no sample data provided, so it is difficult to know how to get data into the system in the right format [e-Science].

SO37. Having an experienced person to assist with initial installation is very good. We suggest having an ICAT mentor for this for future users [SNS].

3.8 IP and Copyright

SO38. All licensing is under a New BSD licence, with no issues yet [e-Science].

SO39. Having the software available as open source is very worthwhile; we welcome the use of the New BSD license [DIAMOND].

4 Recommendations

The following sections cover a list of recommendations for the ICAT team to consider. References to key observations are included, with estimated effort for implementation and perceived importance rated on a scale from 1 to 10 (low to high). The perceived importance is a subjective measure based on the opinion of the assessor from the observations made. Figure 1, page 17, depicts the recommendations graphically on a Perceived Importance/Estimated Effort chart.

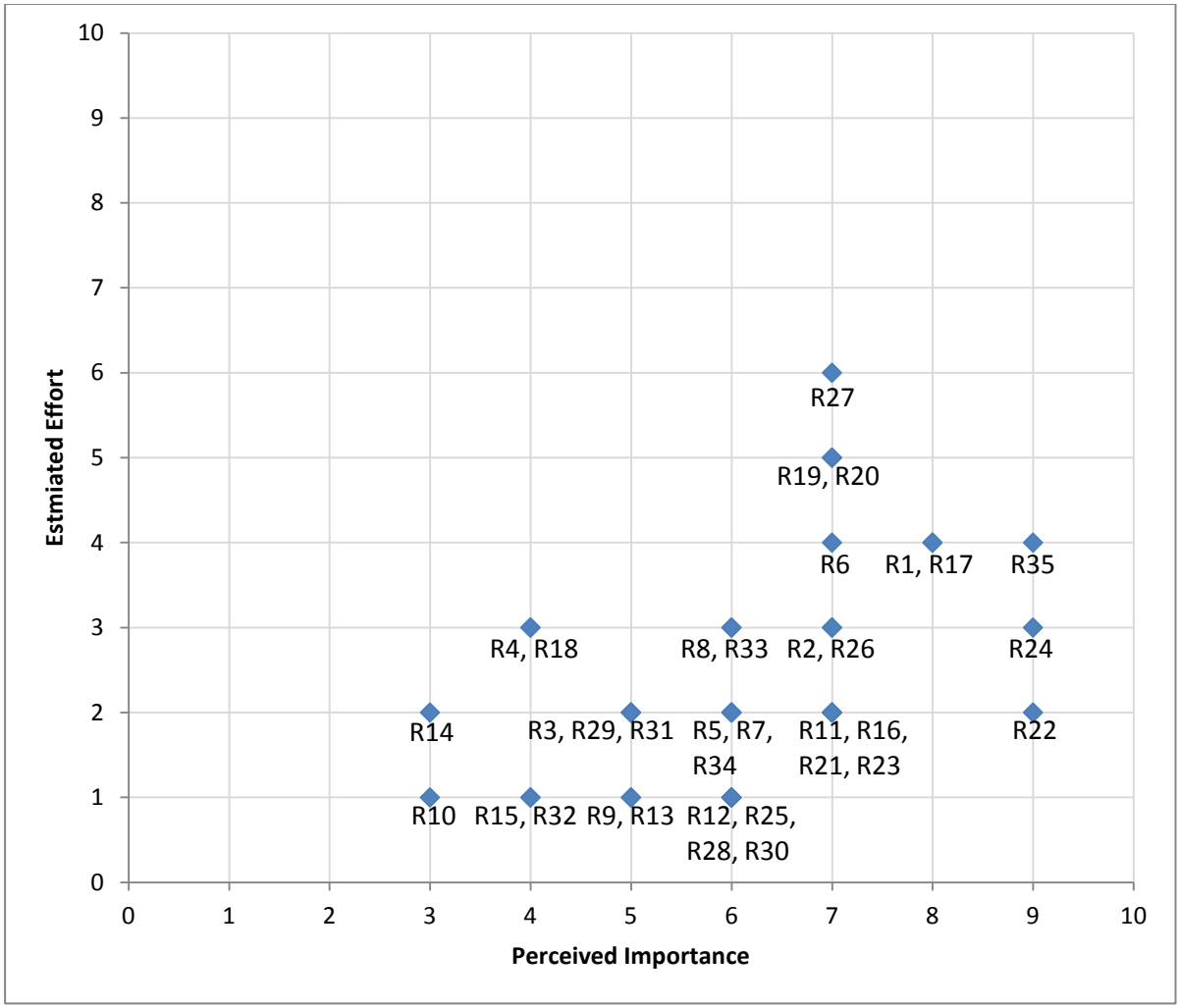

Figure 1: Perceived Importance/Estimated Effort

On this chart, it is clear that there are many recommendations that are perceived as important, and that the bulk of them are clustered around a mid-to-high importance but low-effort point. This is corroborated by the averages of these values being 6.1 for *Perceived Importance* and 2.5 for *Estimated Effort*, with corresponding standard deviations of 1.61 and 1.32 respectively. So the conclusion is that many of the recommendations have a favourable value to cost ratio.

A guiding principle is that the implementation of each recommendation should be as lightweight as possible to achieve the desired effect, while ensuring good use of the resources available. Of course, as the collaboration progresses, each of the recommendations can subsequently be expanded as appropriate.

4.1 Governance

R1. The ICAT team must continue to consolidate their leadership of the collaboration. Their governance structure and strategy needs to be established ideally before collaborators from the PaNdataODI become involved, and should be defined within a clear governance policy. This will clarify the organisational groups and collaborative mechanisms within the governance structure. Input from RAL and PaNdataODI stakeholders should be considered [GO2-6, GO12][Estimated Effort: 4, Perceived Importance: 8].

R2. The creation of a Steering Committee should be considered. As more collaborators become involved, higher-level guidance will inevitably become more important as the complexity of managing the collaboration increases. This organisation would at least include representatives from RAL and PaNdataODI, and perhaps appropriate external experts. Meeting frequency of at least twice a year is recommended [GO12, GO13][Estimated Effort: 3, Perceived Importance: 7].

R3. The manner in which new collaborators become involved in the collaboration should be clearly and transparently defined, in terms of higher-level representation and technical representation. It should build on the ease with which collaborators join currently. This would allow prospective collaborators to interact efficiently with the group and other collaborators at a number of levels in the governance structure [GO11-12, GO13, GO15-16][Estimated Effort: 2, Perceived Importance: 5].

R4. The ICAT team should assume a greater responsibility for organising the Face-to-Face meetings, in at least one previous instance this has been done by ISIS [GO7][Estimated Effort: 3, Perceived Importance: 4].

R5. A 'critical role' / 'bus factor' assessment should be conducted across the collaboration, to determine any areas where a loss of personnel would be disruptive to the ongoing development of the projects [CO6, PO7][Estimated Effort: 2, Perceived Importance: 6].

4.2 Community and Outreach

R6. With Tom Griffin moving towards assuming a customer relationship within the collaboration, replacements for each of his pivotal roles should be considered, including initial technical mentor to external users, product champion and external software tester (see the three specific recommendations below) [CM2][Estimated Effort: 4, Perceived Importance: 7].

R7. The idea of having initial ICAT mentors has been suggested, and this would greatly assist new users technically in adopting the software. This would help to avoid early disenfranchisement with the software, an important consideration with the 11-institution community from PaNdataODI becoming involved. Suitable staff within the ICAT collaboration could be identified to fulfil such a role, with staff allocated responsibility for individual institutions [CM2][SO37][Estimated Effort: 2, Perceived Importance: 6].

R8. Having alpha/beta/release candidate testers who are external to the core development group is of great value, offering a first approach perspective to handling the new releases of the software. Consider canvassing for a manageable number of routine external testers both inside and outside the collaboration to be involved as part of the test process, ensuring their experiences are sufficiently reported [CM2, CM5][Estimated Effort: 3, Perceived Importance: 6].

R9. Once existing or new adopters have become satisfied with their ICAT deployments, promote and disseminate these experiences as success stories to increase confidence in the software and the process of adoption. Identify vehement, quotable ICAT supporters as product champions and raise their profile and publish their success with the software. [CM2, CM6][Estimated Effort: 1, Perceived Importance: 5].

R10. As the size of the collaboration increases, identify opportunities for focused Skype calls (or call groups) to discuss/solve specific technical or organisational issues with a focused set of participants. These should be informal, have a low administrative overhead, have a well-defined mandate and overall be pragmatic. They could report back to the collaboration meetings with progress and final results. [CM4][Estimated Effort: 1, Perceived Important: 3].

R11. The ICAT collaboration, the software, and its successes should be more widely promoted to the broader scientific community to increase the profile of ICAT and potentially to attract strong potential users, and understand what they want from ICAT. This can help to justify future funding and development effort where there is a strong communal need for particular requirements within ICAT. Such an activity could also serve as a net to locate unknown users of the software. With its established publicity mechanisms, the SSI could help directly in this area [CM6-8][Estimated Effort: 2, Perceived Importance: 7].

R12. Outreach to the wider community to request technical feedback on the software, from known as well as currently unknown users, should be considered. This could be conducted in a broad outreach manner, identifying and using key mailing lists, newsletters and other communication channels across the community to solicit these responses, as well as advertising on the project website. This could also be a first step to identifying and bringing unknown users into the collaboration [CM5-CM7][Estimated Effort: 2, Perceived Importance: 6].

4.3 Process

R13. The definition of a concise collaborator induction process, i.e. a very short list of steps and resources, should be considered. This could include a description of how a new collaboration partner organisation should involve itself with the governance group and the technical group. It should also describe getting started with the software and the supporting infrastructure, as well as the methods for communicating with the ICAT team. It should include contact details for requesting help, and reporting difficulties. A new collaboration partner should be able to register his or her participation, and in return receive support from a named mentor within the ICAT team. The mentor should review the requirements of the partner to understand the intended operating environment and use of the software. This should help to avoid incorrect assumptions and misunderstandings [R3, GO11-12, GO13, GO15-16, CM3, PR5][Estimated Effort: 1, Perceived Importance: 5].

R14. The definition of a lightweight conflict resolution process should be considered to resolve differences that may emerge as the collaboration becomes larger. [PR2][Estimated Effort: 2, Perceived Importance: 3].

R15. Generally, processes that are defined and in use should be publicly available, to provide transparency and increase assurance in the management of the collaboration [PR7][Estimated Effort: 1, Perceived Importance: 4].

R16. The various existing methods for ingesting requirements should be streamlined into a single process using a single infrastructure for simplicity and transparency. It would also improve the information on the provenance of requirements, assist in soliciting opinion for requirements, and make it easier to identify common requirements among stakeholders. For example, the issue tracker currently employed on GoogleCode could be a suitable system for this. This would most likely require previous requirements to be entered into the chosen system for completeness [PR8-10, PR12, PR14, PR17][Estimated Effort: 2, Perceived Importance: 7].

R17. It is strongly recommended that the existing community roadmap activity should be expanded and made more inclusive, to ensure all stakeholders, and perhaps the wider community, are involved in this process. Again, this comes back to promotion, management and product assurance. The roadmap should be publicly available, and the

process well-advertised. This will become crucial with the PaNdataODI project [PR7, 10-12, PR15-16, PR18, PR20-24][Estimated Effort: 4, Perceived Importance: 8].

[A method of managing prioritisation across stakeholders was suggested by SNS: Meet every quarter and vote on requirements. Each participant simply has 40 points to spend on the various requirements. At the end of this process, the priority of each of the requirement emerges. However, it is still necessary to deal with popular requirements that cannot be feasibly developed. This process has been running successfully for a couple of years at a facility in the area of materials processing and science.]

R18. The test process used within the ICAT team (mechanisms, technology, staff roles) should be defined, building on the formalisation that has already begun in this area. The results from continuous testing could be published publicly to offer software assurance. It should also be expanded to include external user testing [PR25-PR26][Estimated Effort: 3, Perceived Importance: 4].

R19. The ICAT team should be involved in the definition of facility roll-out procedures for future ICAT versions, in particular in the planning, deployment and testing stages [PR27][Estimated Effort: 5, Perceived Importance: 7].

R20. Progress on standardising and defining the release process, for example around the given six monthly release cycles, should be continued, with an attempt to entice stakeholders into deploying established releases and not bespoke versions. This may not be straightforward, but where possible it would help to avoid informal forking of the code base and enable a release process, a roll-out and a support process which is uniform across all stakeholders [PR28-PR30][Estimated Effort: 5, Perceived Importance: 7].

R21. A technical handover process should be considered for outgoing staff to ensure that vital technical knowledge is retained [CO6][Estimated Effort: 2, Perceived Importance: 7].

4.4 Policies

R22. The definition and use of a deprecation policy, in conjunction with the road-map process, is strongly recommended to avoid important features being phased out without warning in future releases [PO1-PO3][Estimated Effort: 2, Perceived Importance: 9].

R23. The definition of a code contribution policy should be considered to manage any future developments, particularly from PaNdataODI, including guidance on the mechanisms to use (e.g. SVN committer membership and code commits) and how the contributions will be managed [PO4-PO8][Estimated Effort: 2, Perceived Importance: 7].

R24. It is strongly recommended that software release schedules are defined and publicised early to stakeholders and more widely (i.e. defined as a matter of routine based on outputs from the road-mapping process). This should also include changes to the trust store deployed on servers, which is critical to ICAT clients. A minimum of three months' notice should be considered [PO10-11][Estimated Effort: 3, Perceived Importance: 9].

R25. Inclusion of a version change log, perhaps created from the SVN revision history, with commits linked to the issue tracker, for a tagged release, and release notes should be considered [PO12][Estimated Effort: 1, Perceived Importance: 6].

4.5 Product Support

R26. As part of the support response process, the technical expertise of the issue submitter should be taken into account with any technical reply. Where technical expertise is relatively low, this may imply that a more assisted approach is required [PS3][Estimated Effort: 3, Perceived Importance: 7].

R27. This recommendation has been removed from the report.

4.6 Infrastructure

R28. Regular updates to the GoogleCode project front-facing presentation, for example overall software information and benefits, recent developments, links to success stories, resource links, should be considered, to avoid the perception of a stagnant GoogleCode project, which ICAT definitely is not [IN2][SO12][Estimated Effort: 1, Perceived Importance: 6].

R29. Where practical, internal technical, and organisational discussions should be conducted on the public mailing lists to increase transparency and allow stakeholder opinion to be solicited where appropriate [IN3][Estimated Effort: 2, Perceived Importance: 5].

R30. A mailing list should be actively used to announce software releases; an announcement should include information on the version, the release notes, lists of new and deprecated features, lists of known issues, and any relevant organisational changes. This could be done as a matter of routine in a timely manner before a release based on the outputs from the road-mapping activities. It would also provide a starting point for technical discussions and feedback on releases; alternatively these could be conducted within a forum setting e.g. Google Groups [IN3-IN5, IN7][Estimated Effort: 1, Perceived Importance: 6].

R31. A Frequently Asked Questions (FAQ) should be considered to minimise support for already solved common issues. This could be regularly seeded with appropriate solved issues from the issue tracker, perhaps on a release-by-release basis [Estimated Effort: 2, Perceived Importance: 5].

R32. The use of Skype, or similar technology, perhaps with video where helpful, should be considered for appropriate meetings with 2-6 people. It is free, and generally effective and robust for most circumstances involving global collaborators. This could include a joining protocol as part of an established meeting e.g. Skype IDs exchanged in a timely manner beforehand [IN8-IN9][Estimated Effort: 1, Perceived Importance: 4].

R33. The source code repository should be cleaned-up to reflect the simplification improvements made in ICAT 4.0. A clean repository lends itself to software assurance [IN10][Estimated Effort: 3, Perceived Importance: 6].

R34. Information on the collaboration and on the software that is difficult to find, and general opinions on how information is presented, should be identified by asking the stakeholders, and structural improvements made [IN11][Estimated Effort: 2, Perceived Importance: 6].

R35. This recommendation has been removed from the report.

5 Requirements

Below is a summary of specific requirements or upcoming areas for requirements from stakeholders. These have emerged out of the assessment, and may be useful for the ICAT team to consider (should they not already be aware of them).

ISIS:

REQ1. ISIS is keen to get involved with derived data, to enable scientists to upload derived data to the ICAT instance. This may entail development work in TopCAT. ISIS anticipates a requirement for ICAT to support non-experimental data such as data usage statistics. At the moment, ICAT can record this information but it is not searchable.

REQ2. There are some challenges for e.g. ontologies, searching across many catalogues.

DIAMOND:

REQ3. The original priority was to ensure data would not be lost. Now that this has been solved, there is a need to satisfy greater requirements.

REQ4. They have developed an Eclipse RCP-based tool which incorporates ICAT, which enables users to log in against the API, to list experiment sessions after login, and to drill down to file data locations, similar to TopCAT. ICAT is now within the users' daily operating environment for their data, and this tool can be used offsite for collaboration. During an experiment, users can observe files being generated in real time. Metadata for this use case will grow – this will inform future database requirements.

REQ5. New requirements will be driven from use cases within DIAMOND, as well as the DOI use case for publishing. The question remains about how far people want to use this within a LIMS, or within a publisher system.

CLF:

REQ6. As ICAT was originally designed for experimental data, using it for other data such as diagnostic data, leads to new requirements; a careful assessment is required before applying ICAT to other domains.

REQ7. This requirement has been removed from the report.

SNS:

REQ8. A lot of new requirements are coming in, e.g. how do I get data across facilities as a user? At the moment, users need to log in twice (e.g. within the MANTID project). To achieve this, working now on consolidation of data across different facilities is a big thing. The key is Single Sign-On (SSO), but not sure how quickly ORNL can get into this. Currently use proprietary facility-based authorisation. But this would be very beneficial for users, especially neutron users who want to share data across different facilities.

REQ9. SNS would like to have the ICAT API support the provision of certain unprivileged metadata without the need for authentication and authorization. In particular SNS wish that ICAT return the file locations for a particular instrument and run. The actual authentication and authorisation would be done later at file access time. This would bring in much more traffic to ICAT, and would be used widely used.

REQ10. One of the biggest tasks is to accomplish automated data reduction. This requires a running ICAT instance, associated with a catalogue of raw experimental data. Then, perhaps from an instrument's metadata, you pull out the specific configuration for this cycle and run the data reduction. This is currently done manually, and doing this automatically would be useful for the users, and would enable them to do analysis and

plotting, and work towards publication more quickly. ICAT plays a very important role at SNS in making automatic data reduction work, and the testing of the system looks good. SNS hopes to expand on this work.

REQ11. SNS demonstrated TopCAT to someone interested in it. There are lots of pop-up windows; could these be reduced? Could an HCI specialist have to look at it?

REQ12. It is currently difficult to go instrument to instrument to consolidate shortlist of parameters for ICAT so that it will be useful for their data analysis if they want to redo data reduction. Would it be possible to e.g. ask ICAT to return all the data I own within a temperate range? Requests for this type of requirement are expected to increase.

ELLETRA:

REQ13. By June, when PaNdataODI deliverable D4.1 will be ready, the survey should create a features wish-list for the data catalogue. This essentially would be a wish-list based on ICAT 4, taking into account requirements from at least the x-ray and neutron sources.

REQ14. One definite thing that will be asked for is a feature to enable cross-facility searching.

REQ15. The core of PaNdata is data policy, which extends to software as well. Having features to enforce data policies may be useful.

6 Conclusions

From the 15 conducted interviews, the assessment has yielded a total of 92 observations on the collaboration, 39 observations on the software and 15 discussed requirements. The subsequent analysis of the collaborative observations has led to a series of 33 recommendations for consideration by the ICAT team.

A number of important observations were made. Firstly, that ICAT has become an integral part of the operational processes of the stakeholders, for whom there is now a clear dependency on the software [CO1]. The collaborative processes and mechanisms in place have so far catered well for the needs of the collaboration [CO2], although it is clearly important that these existing mechanisms are appropriately expanded upon to ensure, and provide assurance for, a sustainable future for the collaboration [CO3, CO4]. In addition, the ICAT team is very much aware that the definition of more governance, process and policy are required to manage the increased uptake that is expected [CO5, GO2, GO3, GO5, GO6].

The set of observations logically led to a number of key recommendations. As well as a definition of clearly defined governance [R1], a number of process and policy improvements should be considered. In terms of process, a well-defined induction process for new collaborators is recommended [R13], as well as a more streamlined and efficient requirements ingest process [R16]. Of particular importance is the expansion and promotion of the road-mapping activity [R17] which will become crucial as more stakeholders such as PaNdataODI, and their requirements, become part of the collaboration. To support these processes, and to further promote smooth integration with new collaborators, serious consideration should be given to define deprecation, code contribution and release policies [R22, R23, R24]. To ensure a higher profile and level of assurance for the ICAT software and collaboration, and to increase uptake in existing and perhaps unchartered areas, effort should be considered for promoting these aspects more widely [R9, R11]. To increase community involvement and integration with stakeholders, external product testers [R18], ICAT team involvement with facility roll-out plans [R19] and an ICAT software mentoring capability [R7] should also be considered.