Report on Usefulness Assessment

For ESIP Products & Services Fast Track Proposal

* Problem Statement
* Goal of Research
* Methods
  + Literature Review
    - Usability literature
    - Technology assessment literature related to NASA TRLs
  + Use Cases
  + Ranking Effectiveness (“Usefulness”) of Communication Vehicles
  + Development of Evaluator Checklists
  + Feedback on draft Evaluator Checklists
  + Draft Guidelines in support of user interface usability assessment
* Discussion
* Recommendations / Next steps

**Problem Statement:**

One of the strengths of the ESIP Federation is the availability of opportunities to probe the myriad problems associated with the stewardship of Earth and space data, and offer possible solutions in a collaborative, research-oriented environment. These explorations usually include some kind of community evaluation – either formal or informal. One of the main venues within ESIP in which exploration of this kind occurs is within the ESIP Products & Services (P&S) Committee’s Testbeds. While the original P&S Testbed is focused upon the first phase of research and/or product development, considered the “incubation” phase, more recently, additional testbeds associated with the P&S have been focused upon later stages of research and product development in which the successful incubator projects move toward “infusion” into more established, sustained and financially supportable environments. Projects at this infusion stage that ESIP is helping to evaluate are those of NASA’s Advanced Information Systems Technology (AIST). See: <https://esto.nasa.gov/info_technologies_aist.html>.

With all of these projects, community evaluation is an important aspect of the P&S’s work. To date, however, practical heuristics have not been identified by which evaluators can assess the overall effectiveness i.e., usefulness of the means by which a research project conveys information about itself, such as the research problem being studied and its findings and/or proposed solution. For instance, how well can another researcher who has a problem similar to the one being explored determine the specific use case(s) being addressed by the project? How easy is it to navigate the web pages of the research project in order to find the information that s/he needs to decide whether the research product is promising for his/her own use? These kinds of questions are asked by ESIP evaluators, but also by ESIP members (and anyone perusing the information). Rules of thumb that can be applied by research PIs, product developers as well as evaluators to check that communication vehicles being created are effectively conveying the most important information would be generally useful to the ESIP community.

**Goal of Research:**

To develop a heuristic[[1]](#footnote-1) approach to evaluating the effectiveness / usefulness of the various means of communication by which a research project informs one or more targeted audiences about the project’s goals, objectives, technologies, methods, workflow or any other information pertinent to those interested in adopting or adapting them.

**Methods:**

**1. Literature Review / Annotated Bibliography**

* **Usability Literature**

**Nielsen, Jakob, “Usability 101: “Introduction to usability”, Nielsen Norman Group, (January 4, 2012), last retrieved from** [**https://www.nngroup.com/articles/usability-101-introduction-to-usability/**](https://www.nngroup.com/articles/usability-101-introduction-to-usability/) **on 9 February 2017.**

This is a seminal article for the web-based usability literature that provides an easy to understand definition of what “usability” is, and the five quality components that can be used to assess usability (learnability, efficiency, memorability, errors, and satisfaction). These five components are included in the Evaluator Checklist format created for this project. This article also described the relationship between “utility” and “usability” that should result in a “useful” website. This simple, but powerful concept was used to illustrate the approach that was taken for assessing the communication vehicles associated with a research project and for creating the evaluation heuristics.

**Mangal, Nimit, “What is usability?”, (2013, September 19), last retrieved from** [**https://nimitmangal.wordpress.com/2013/09/19/what-is-usability/**](https://nimitmangal.wordpress.com/2013/09/19/what-is-usability/) **on 9 February 2017.**

This article is from the blog of a web developer and consultant that contains a useful image of the relationship between “utility” and “usability” that should result in a “useful” website as described by Jakob Nielsen in his “Usability 101” article referenced above. The image was used in the various presentations of the topic including the presentation at ESIP Winter 2017 (attached as a PDF).

**UX Mastery. “UX Techniques”, (N.D.), last retrieved from** [**http://uxmastery.com/resources/techniques/**](http://uxmastery.com/resources/techniques/) **on 9 February 2017.**

This article is one of many resources on the UXmastery.com website that provides resources for those user experience and user design professionals from beginner to advanced. This particular article provided a good list of techniques for assessing usability, with definitions, and explanations for when and how to use the techniques. For purposes of this project, the most useful techniques discussed in this article were the “heuristic reviews” and “use cases”.

**UX for the Masses. “A guide to carrying out usability reviews”, (February 11, 2011), last retrieved from** [**http://www.uxforthemasses.com/usability-reviews/**](http://www.uxforthemasses.com/usability-reviews/) **9 February 2017.**

This article provided an easy to understand way to conduct usability reviews that used as examples the “heuristic review”, and the “cognitive walkthrough”, both of which seemed particularly applicable to this project (in addition to the employment of “use cases”). It also included a link to a “Usability Review Template” in Excel format that is included as a tab called “Usability Guidelines” on the attached Excel file containing the Evaluator Checklists.

**Usability Professionals Association. “Cognitive Walkthrough”, published on Usability Body of Knowledge (N.D.), last retrieved from** [**http://usabilitybok.org/print/cognitive-walkthrough**](http://usabilitybok.org/print/cognitive-walkthrough) **on 9 February 2017.**

This article provided a very helpful, and easily understandable explanation of the cognitive walkthrough technique, its benefits, advantages and disadvantages, and how to use it. The “streamlined approach” discussed within the article was particularly applicable for this project. The article also contained related links that are promising for those wishing to learn more about this technique.

**Usability Professionals Association. “Heuristic Evaluation”, published on Usability Body of Knowledge (N.D.). Last retrieved from** [**http://usabilitybok.org/print/heuristic-evaluation**](http://usabilitybok.org/print/heuristic-evaluation) **on 9 February 2017.**

This article provided a very helpful, easily understandable, but rather extensive explanation of the heuristic evaluation technique, its benefits, advantages and disadvantages, and how to use it. The technique is tied closely with the work done by Jakob Nielsen, previously cited. The article also contained related links, references, and more formally published resources that are promising for those wishing to learn more about this technique.

* **Technology Assessment Literature (focusing on NASA’s TRL levels)**

**Zelkowitz, Marvin. V., Software Engineering Technology Infusion within NASA, I*EEE Transactions on Engineering Management*, 43.3 (1996): 250-61. PDF:** [**https://www.cs.umd.edu/~mvz/handouts/eng.mgmt.pdf**](https://www.cs.umd.edu/~mvz/handouts/eng.mgmt.pdf)  
 This article was seminal for understanding the importance of the technology ***infusion*** process to the broader process of change in the current state of technology practices, using software engineering technologies within NASA as examples. The article was published in 1996, and therefore, very early in the days of the use of the World Wide Web, so both the example technologies, and the discussions of communication models are dated, albeit historically interesting. Nevertheless, the distinction that Zelkowitz makes between “technology transfer” (defined as “the adoption of a new method by large segments of an industry” which had been the more published topic) and “technology infusion” (defined as “the adoption of a new technology by an individual organization”) was very helpful for understanding terms that often seemed to be used interchangeably, although not necessarily in the NASA context. In addition, Zelkowitz’ discussion of the Quality Improvement Paradigm (QIP) used by NASA’s Software Engineering Laboratory (SEL) provided the means to think about an approach that, abstracted, could be applied to anyone investigating a research project or new technology for purposes of assessment or infusion / adoption. The three steps of the QIP seem simple, but clear, i.e., “understanding” the research project or technology, “assessing” it, and standing back to look at the “packaging” of the research project or technology in terms of the documentation or other communication vehicles produced (p.16). As described by Zelkowitz, the steps are iterative which is important to remember when using the Evaluator Checklist developed as part of this project. See more about these timing issues in the Development of Assessment Tool section (4) below. The conclusions within this article are also interesting for those focused upon more specifically upon software engineering practice and technology assessment.

**Olechowski, Alison L, and Eppinger, Steven D., and Joglekar, Nitkin, “Technology Readiness Levels at 40: A Study of State-of-the-Art use, Challenges, and Opportunities” (April 1, 2015). MIT Sloan Research Paper No. 5127-15. Available at SSRN:** [**http://ssrn.com/abstract=2588524**](http://ssrn.com/abstract=2588524) **or** [**http://dx.doi.org/10.2139/ssrn.2588524**](http://dx.doi.org/10.2139/ssrn.2588524) **.**   
 This article was very helpful for providing a more updated view than the descriptions in the Zelkowitz article of how the TRL process has been applied within NASA. The article offered a mapping of TRL levels to system development lifecycle steps that provided a familiar comparison. The article contained findings about the use and effectiveness of the TRL methodology based on interviews with those who had experience applying them from a number of organizations (including NASA), and from an extensive literature review. The discussion of which and how other organizations and agencies have used the TRL methodology was helpful to get a better understanding of how the TRL methodology has been or could be adapted. The list of references from the literature review was very helpful in getting a good, academic understanding of the TRL methodology.

**Mankins, John C., “Approaches to Strategic Research and Technology (R&T) Analysis and Road Mapping.” *Acta Astronautica,* 51:1-9 (January 2002): 3-21.** In this article, Mankins offers a good overview of the history of NASA’s Technology Readiness Level (TRL) approach to technology management at all levels in an organization. In addition to definitions, the article discusses how the TRLs were designed to provide information on the current and desired level of maturity of a technology for a particular application. In addition, there is an interesting discussion about how the developers of an application, or research project for purposes of this discussion, can pull together a “***road map***” that can guide the advocacy, planning, implementation, and assessment of the technology, and significantly improve the chances for the successful infusion or adoption of the technology / research by others. The road map, a collection of concepts and documents, can be used as a strategic tool for describing the technological “place” to which the research project and /or technology is headed. Ideally, the strategic road map should be discernible from the communication vehicles that the research project and/or technology uses to inform others of its goals and objectives (p.13).

**Kostoff, Ronald N., and Schaller, Robert R., “Science and Technology Roadmaps.” *IEEE Transactions on Engineering Management,* 45:2 (May 2001).** This article discusses the use of road maps in detail, especially for purposes of enhancing the efficiency of the technology transfer process within science and technology arenas. Road maps are described as being especially useful as decision aids to improve the coordination of activities and resources in complex and uncertain environments. One of the specific uses for which road maps are recommended is for enhancing communication among stakeholders of a research project or technology, such as researchers, technologies, product managers, suppliers, users, and others. Kostoff and Schaller offer common definitions for and characteristics of different types of road maps as well as a detailed explanation of the process for constructing high-quality road maps.

**Comstock, Douglas A., “Technology Development and Infusion from NASA’s Innovative Partnerships Program.” IEEEAC paper #1538, Version 3. (updated January 3, 2008).** [**https://www.nasa.gov/pdf/324711main\_ieeeac\_paper\_138.pdf**](https://www.nasa.gov/pdf/324711main_ieeeac_paper_138.pdf)Although NASA’s TRLs are not explicitly discussed within this article, Comstock does provide an organizational context and discussion of the challenges and obstacles for technology infusion within an organization, illustrated by examples from within NASA’s Innovative Partnerships program. While this context is interesting, the emphasis made upon the need for *communication* about how the technology or research could be used or adapted by others during the development lifecycle was particularly helpful given the goals of this research. In addition, the discussion of best practices related to communication for purposes of technology infusion was particularly helpful, as were both the categories of questions that potential infusion decisionmakers usually ask (i.e., those related to performance, schedule, cost, and risk), and the examples of questions within each of the categories. (p. 9 - 10).

**2. Use Cases:**The following Use Cases were developed to help scope the research:

* Use Case 1: Research Domain Expert
  + I am a domain expert in tropical cyclones and am looking for web-based tools or services to help my research group with certain types of analysis. I am interested in guidance on how to determine if a tool or service will be useful for addressing the research needs of my group.
* Use Case 2: Workflow Domain Expert
  + I am a domain expert in tropical cyclones looking for web-based tools or services to help my research group perform certain types of analysis more efficiently and productively without requiring a lot of hands-on training or long ramp-up time.
* Use Case 3: Tool-builder
  + I am a researcher building a tool to provide outcomes from model analyses for tropical cyclone research. I want this tool to be adopted by my research community. On what kinds of factors or points about my tool should I focus to help me meet that goal?
* Use Case 4: AIST Evaluator
  + I am a researcher building a tool to provide outcomes from model analyses for tropical cyclone research. I want this tool to be adopted by my research community. On what kinds of factors or points about my tool should I focus to help me meet that goal?

Note that in the course of the research, the Products & Services Committee Chair, and ESIP Staff advised that the use case of most relevance and immediate need was for the Research Domain Expert. As a result, the Evaluator Checklist has been most fully developed for this use case. The others have basic questions associated with them, but definitely need more community input and testing.

**3. Ranking Effectiveness (“Usefulness”) of Communication Vehicles (CVs)**In Jakob Nielsen’s article, “Usability 101: Introduction to Usability”, he discusses five quality components of usability which have been used as the categories for ranking the effectiveness of the communication vehicles that a research project, tool or technology uses to convey information about itself. While Nielsen describes these components as being important for assessing the “usability” of a user interface, if they are applied in order to determine the overall “utility” (defined as the “functionality” associated with a CV) that is offered by all the CVs available, the components seem very appropriate for ranking the effectiveness/ usefulness of the CVs per the definition of the term as described above (utility + usability = usefulness).

The five components can be better understood by asking the following questions:

* **Learnability:** How easy is it for users to accomplish basic tasks the first time they encounter the CVs?
* **Efficiency:**  Once users have learned the design, how quickly can they answer the questions they have or the tasks that they wish to accomplish?
* **Memorability:** When users return to the CVs after a period of not using them, how easily can they establish proficiency, or (re)find what they are looking for?
* **Errors:** How many errors do users make that result from the CV, how severe are the errors, and how easily can users recover from the errors? (Note a discussion of “errors” in user interface design can be found at: <https://www.nngroup.com/articles/slips/> )?
* **Satisfaction:** How pleasant is it to use the CVs?

Questions that are specific to the points of view represented by the experts defined in the use cases can be found on the Evaluator Checklist for each of these categories.

**4. Development of Assessment Tool as Evaluator Checklist**The attached assessment tool (in MS Excel format) for the evaluators was derived from the application of the concepts found to be pertinent in the usability literature to the technology infusion & adoption concepts in the TRL evaluation literature. It has been built to accommodate the perspectives of the four experts described in the use cases, but, hopefully, is broad enough to be applicable to other use cases as well which may include non-experts.

The Evaluator Checklist for the Research Domain Expert is the most comprehensive in terms of the specific questions that are included within the categories of “understanding”, “assessing” and “packaging” per the scope of the research. Each of the checklists has similar questions at present, but could be much more tailored to the specific questions and/or tasks that each expert might have (as well as a non-expert). In addition, further narrowing the use case to include one or more “personas” (defined as a fictitious identify that reflects one of the user groups for whom the CV is designed) might be productive. As previously mentioned, the questions focus primarily upon CVs that may include project websites, documentation, lists of references, online help text and/or tutorials, testimonials, etc. The questions could be used at different points in the evaluation process. For example, the questions related to “understanding” the research project, tool or technology could and probably should be asked, answered and scored early in the evaluation process as it would be more difficult to accurately remember first impressions later in the evaluation process.

The scoring for the questions in the five component categories was designed to be qualitative rather than quantitative in order to lessen the sense of “grading” a communication vehicle, and thereby implying failure. Given the fact that this kind of evaluation is, by definition, subjective, the qualitative scoring (Very poor / Poor / Moderate / Good / Excellent) plus the opportunity for the evaluator to have ample room for comments about the scoring and/or suggestions was intended to convey constructive information that could be used to improve the CVs.

**5. Feedback on draft Evaluator Checklist for the Domain Research Expert**The Research Domain Expert Evaluator Checklist was initially presented for community feedback at the ESIP Winter 2017 meeting in Washington, DC in January of 2017 at a breakout session. Comments after the presentation (attached as a PDF file) were as follows:

* After generating questions from the ESIP community (and others, as feasible) for the other domain experts, it would be very useful to do some testing of this instrument on a research project or other kind of project with at least the four domain experts that are described within the use cases, if not groups of domain experts and then correlate the responses to see how the evaluation process categories and the ranking categories work with each of the groups. Part of the purpose for this testing would be to see whether the matrix approach is too subjective to be useful.
* Would be useful to know more about how the different types of assessments work (i.e., a usability assessment in the classic case of user interface design, and a TRL assessment), and have more discussion about how the application of one to the other could be done.
* Would be useful to know more about how TRLs have been adopted and adapted by the Earth Cube community and other communities as well. (See Olechowskie, et al, article referenced above which provides some information on this topic as well as other articles & references noted in the article).
* Once the instrument has been tested, more specific guidelines would be very helpful, with examples from actual websites or other communication vehicles.

It is useful to note, that several people at the session were interested in using / adapting the evaluation heuristics for they found them potentially helpful to assess project web sites that they were planning to undertake, and/or training that they were planning to provide to early career scientists. The draft Research Domain Expert Checklist was sent to one of them (Stace Beaulieu at BCO-DMO), and has been requested by a representative for the Open Science Framework).

**6. Draft Guidelines in support of the user interface usability assessment aspects of the Evaluator Checklist**For those not familiar with the process of testing the usability of user interfaces to CVs such as those of research project web sites, it might be difficult to know what to look for when stepping back to evaluate what is seen on the web page in terms of its presentation of information rather than for its intellectual content. For this reason, “Usability Guidelines” have been added to the attached Evaluation Checklist Excel file that are included as part of the Usability Review Template link in the article from UX for the Masses referenced above, called “A guide to carrying out usability reviews” (February 11, 2011). The areas on a web site or other CV to which one using the Usability Guidelines is directed include its form and functionality, its homepage / starting page; its navigation, its search capabilities; its control and feedback, its forms, its errors; its content and text; its help and its performance. The Usability Guidelines include more specific questions within each of these areas and rates them by importance to the successful user experience.

**Discussion:**Evaluation process: The questions that are listed on the Evaluator Checklist relate to the information that the domain expert could / would ask in order to achieve the evaluation of the research project, technology or product itself. In order to actually do the scoring of the CVs according to the usability categories, however, it would be necessary for the evaluator to follow a three-step process for each question. The first step would be to look at the intellectual content found on the CV that is available to answer the question while also noting where the information is found. The second step of the process is to assess the features of the CV where the information is found, and score those features from Very poor – Excellent. The third step would be to add comments that explain what aspects of the feature were particularly helpful, or could be improved. The final step would be an important one for each evaluator to do in order to provide the most constructive information to the creators of the CVs.

For those not familiar with user interface testing, it may be difficult to distinguish and identify the features and/or capabilities seen on a CV. For this reason, the Usability Guidelines have been included on the worksheet that contains the Evaluator Checklists. As mentioned above, the categories that are included on the Usability Guidelines include the features and functionality of the CV, the homepage or starting page, its navigation, its search function, the means of control and feedback, its forms, its content and text, the help available, and the overall performance. The questions asked under each of these categories in the Usability Guidelines can help an evaluator stand back from the intellectual content of the CV to think about and evaluate whether and how the semantic information came through.

For example, let’s say an evaluator is just beginning to look at NASA’s Hurricanes and Tropical Storms website at: <https://www.nasa.gov/mission_pages/hurricanes/main/index.html> in order to answer the following question in the Understanding part of the evaluation process: “What is / are the research domain(s) with which this research is associated?” Presumably, the first place to look for the information that will allow the evaluator to answer this question would be the home page. The Usability Guidelines note that a feature of high importance for the homepage or starting page is whether it is “effective in orienting and directing users to their desired information and tasks. Users should be able to work out where they need to go to complete a given task.” In this case, the task would be to answer the question about what research domains are associated with this website. On the homepage of this website, below the section of the home page that identifies it as being about “Hurricanes and Tropical Storms”, there is an arrow pointing to subtopics that include an “Overview”. Clicking on the Overview tab takes the evaluator to another page that provides subtopics related to hurricanes with small images, a single sentence explanation and a link to subpages. What is not immediately obvious is what or whether there are subtopics related to tropical storms and/or what the semantic relationship is between hurricanes and tropical storms. Is one topic a subset of the other? To a research domain expert, that may not be a question worth answering on this page of the website; to a member of the general public, it might be a relevant question. In either case, it would be necessary to investigate further within the site to see what or whether there were more specific topics covered that relate to tropical storms. If the evaluator were to backtrack to the home page, there is a section on the lefthand side that includes links to “Related Topics.” Tropical Storms are not listed here either, so further navigation is necessary. Given these features, how should / would an evaluator score the “learnability” of the website, i.e., how easy was it to learn the answer to the question about what research domains are associated with this website?

Clearly, an evaluator could spend a great deal of time looking closely at all the features of a CV, and also answering the questions posed at each step of the evaluation process. The amount of time allocated will have to be determined based on the role of the evaluator, and the importance of that role to the overall evaluation process. What might be more feasible would be to have two stages of evaluation (or two different evaluators?), one to answer the technical questions related to the technology infusion or adoptability potential for a research project or product, and another to address the effectiveness of the CVs in communicating important information that would influence its technology infusion or adoptability potential.

**Recommendations/ Next steps:**

* Add to questions for 3 use cases besides the research domain expert
* Add different “persona” approach to augment and clarify the Use Cases
* Test in a couple of venues with web sites relating to each of the use cases, i.e., a research project, a tool site, a project describing a workflow (hard to find?), an AIST technology project, e.g., – e.g., with Stace Beaulieu, from BCO-DMO, and/or Natalie Meyers from the Open Science Framework on research project web sites. Test for feedback upon, for example:
  + The “Usefulness” approach – melding of the two areas of research
  + The viability of the questions from each use case / persona POV
  + The utility of the rating mechanism (i.e., as subjective, but non-quantitative)
* Iterate the questions on the Checklists based on the testing
* Re-draft the guidelines based on the testing using a specific research web site as an example
* Perhaps write a paper for publication after testing in conjunction with the AIST assessment if these heuristics prove useful to those efforts

1. An easily understood and seemingly quite apt definition of a “heuristic approach” from Wikipedia states that “A **heuristic** technique (/hjᵿˈrɪstᵻk/; Ancient Greek: εὑρίσκω, "find" or "discover"), often called simply a **heuristic**, is any **approach** to problem solving, learning, or discovery that employs a practical method not guaranteed to be optimal or perfect, but sufficient for the immediate goals.” (<https://en.wikipedia.org/wiki/Heuristic> ). [↑](#footnote-ref-1)