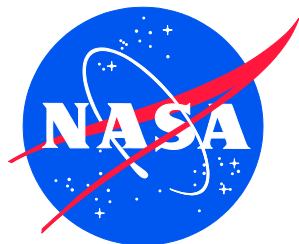


NASA/TM-2015-218991/Volume I
NESC-RP-14-00950



International Space Station (ISS) Anomalies Trending Study

*Robert J. Beil/NESC and Timothy K. Brady/NESC
Langley Research Center, Hampton, Virginia*

*Delmar C. Foster
Data Mining USA, Kennedy Space Center, Florida*

*Robert R. Graber
Science Applications International Corporation, Houston, Texas*

*Jane T. Malin
Johnson Space Center, Houston, Texas*

*Carroll G. Thornesbery
S&K Aerospace, Houston, Texas*

*David R. Throop
Jacobs Technology, Inc., Houston, Texas*

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.

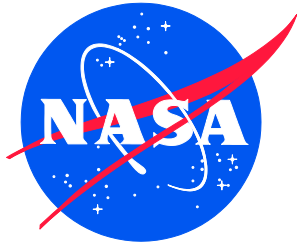
- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>
- E-mail your question to help@sti.nasa.gov
- Phone the NASA STI Information Desk at 757-864-9658
- Write to:
NASA STI Information Desk
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

NASA/TM-2015-218991/Volume I
NESC-RP-14-00950



International Space Station (ISS) Anomalies Trending Study

*Robert J. Beil/NESC and Timothy K. Brady/NESC
Langley Research Center, Hampton, Virginia*

*Delmar C. Foster
Data Mining USA, Kennedy Space Center, Florida*

*Robert R. Graber
Science Applications International Corporation, Houston, Texas*

*Jane T. Malin
Johnson Space Center, Houston, Texas*

*Carroll G. Thornesbery
S&K Aerospace, Houston, Texas*

*David R. Throop
Jacobs Technology, Inc., Houston, Texas*

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

December 2015

The use of trademarks or names of manufacturers in the report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

	<p align="center">NASA Engineering and Safety Center Technical Assessment Report</p>	<p>Document #: NESC-RP- 14-00950</p>	<p>Version: 1.0</p>
<p>Title: ISS Anomalies Trending Study</p>		<p>Page #: 1 of 48</p>	

International Space Station (ISS) Anomalies Trending Study

September 24, 2015

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		2 of 48	

Report Approval and Revision History

NOTE: This document was approved at the September 24, 2015, NRB. This document was submitted to the NESC Director on October 5, 2015, for configuration control.

Approved: _____	<i>Original Signature on File</i>	10/5/15
	NESC Director	Date

Version	Description of Revision	Office of Primary Responsibility	Effective Date
1.0	Initial Release	Robert J. Beil, NESC Systems Engineering Office (SEO), KSC	9/24/15

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		ISS Anomalies Trending Study	
		Page #: 3 of 48	

Table of Contents

Technical Assessment Report	5
1.0 Notification and Authorization	5
2.0 Signature Page.....	6
3.0 Team List	7
4.0 Executive Summary	8
5.0 Assessment Plan	12
6.0 Description of Data Sub-team Tasks.....	12
6.1 Team Methodology	12
6.2 Data Sources	14
6.2.1 Anomaly and Problem Reporting Data Sources	14
6.2.2 Additional Data Sources	14
6.3 Data Extract, Transform, and Load (ETL).....	15
6.3.1 Extract	15
6.3.2 Transform.....	18
6.3.3 Load	19
6.4 Tool Suite.....	20
6.4.1 Search	20
6.4.2 Data Mining to Enhance Search	21
6.4.3 Data Visualization.....	25
6.5 Products Used, Purchased, and/or Developed	29
6.5.1 Data Sets and Data Set Documentation	29
6.5.2 Software and Software Reference Documentation	30
6.5.3 Guides and Training Products.....	31
7.0 Analysis Results.....	32
7.1 Results of Discipline Analysis	32
7.2 Data Enrichment Results	35
7.3 Topic of Interest.....	35
7.3.1 Relating System Hazards and Causes with Problem or Anomaly Occurrences	35
7.4 Description of Future Analysis Plans	42
8.0 Findings, Observations, and NESC Recommendations.....	42
8.1 Findings	42
8.2 Observations	43
8.3 NESC Recommendations	44
9.0 Alternate Viewpoint.....	45
10.0 Other Deliverables	45
11.0 Lessons Learned.....	45
11.1 Preventing Errors in Problem Reporting Codes.....	45
11.1.1 Description.....	45
11.1.2 Corrective and/or Preventive Actions	46

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		ISS Anomalies Trending Study	
		Page #: 4 of 48	

12.0	Recommendations for NASA Standards and Specifications.....	46
13.0	Definition of Terms.....	46
14.0	Acronym List.....	47
15.0	References.....	48
16.0	Appendices (separate volume)	48

List of Figures

Figure 6.1-1.	General Interaction with Discipline Experts to Support Analysis of ISS Anomalies	13
Figure 6.3-1.	ISS Data Sets Extraction, Transformation, and Load	15
Figure 6.3.2-1.	Transformed Fields Examples	18
Figure 6.3.2-2.	Field Addition for Record Integrity Example	19
Figure 6.4.2-1.	View of NESC Data Subteam Activities	21
Figure 6.4.3-1.	Data Visualization Dashboard	27
Figure 6.4.3-2.	Results of a Flamenco+ Keyword Search for “Joint”	28
Figure 7.1-1.	Trends of “ISS Computers” Failures from 2009 to 2014.....	33
Figure 7.1-2.	Nonconformances Containing “Smoke” or “Fire” and “Alarm”	34
Figure 7.3.1-1.	NASA ISS Hazard Data System Search Page	37
Figure 7.3.1-2.	Tableau® Search Screen with Three Search Parameters	40
Figure 7.3.1-3.	Anomaly Text Information Results for Associated Hazard Components/Items.....	41
Figure 7.3.1-4.	Failure Mode Descriptions Associated with Identified Anomaly Records.....	41
Figure 7.3.1-5.	Part Numbers and Descriptions Associated with Identified Anomaly Records	42

List of Tables

Table 6.2-1.	Ancillary Data Sources	14
Table 6.3.1-1.	Data Extraction Date for Each Record System.....	16
Table 6.3.1-2.	Structured and Unstructured Fields Used in Merged Data Set	17
Table 6.3.1-3.	SCR and MADS Data Fields	18
Table 6.3.2-1.	Transformed Fields	19

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP- 14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 5 of 48	

Technical Assessment Report

1.0 Notification and Authorization

The NASA Engineering and Safety Center (NESC) set out to utilize data mining and trending techniques to review the anomaly history of the International Space Station (ISS) and provide tools for discipline experts not involved with the ISS Program to search anomaly data to aid in identification of areas that may warrant further investigation. Additionally, the assessment team aimed to develop an approach and skillset for integrating data sets, with the intent of providing an enriched data set for discipline experts to investigate that is easier to navigate, particularly in light of ISS aging and the plan to extend its life into the late 2020s.

Mr. Robert Beil, NESC Systems Engineering Office (SEO), NASA Kennedy Space Center (KSC), was selected to lead this assessment. The key stakeholders for this assessment were Mr. Timmy Wilson, Director, NESC, and Mr. Michael Suffredini, Manager, ISS Program Office.

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 6 of 48	

2.0 Signature Page

Submitted by:

Team Signature Page on File – 10/30/15

Mr. Robert J. Beil Date

Significant Contributors:

Mr. Timothy K. Brady Date

Mr. Delmar C. Foster Date

Mr. Robert R. Graber Date

Ms. Jane T. Malin Date

Mr. Carroll G. Thronesbery Date


Mr. David R. Throop Date

Signatories declare the findings, observations, and NESC recommendations compiled in the report are factually based from data extracted from program/project documents, contractor reports, and open literature, and/or generated from independently conducted tests, analyses, and inspections.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		7 of 48	

3.0 Team List

Name	Discipline	Organization
Core Team		
Bob Beil	NESC Lead	KSC
Tim Brady	NESC Deputy Lead	JSC
Linda Moore	MTSO Program Analyst	LaRC
Land Fleming	Data Mining, Flamenco+ Customization	Jacobs, JSC
Delmar Foster	Data Mining, SAS®, Tableau®	Data Mining USA, KSC
Jane Malin	Data Mining, Use Case Design, Vetting	JSC
Ali Shaykhian	Database and Information Technology Support	KSC
Carroll Thronesbery	User Interface, Metrics	SKA, JSC
David Throop	STAT Customization, Mining, and Integration	Jacobs, JSC
Consultants		
Bob Graber	Data Consultant	SAIC, JSC
Dave Hamilton	Technical Expert	JSC
Administrative Support		
Linda Burgess	Planning and Control Analyst	LaRC/AMA
Jonay Campbell	Technical Writer	LaRC/NG
Diane Sarrazin	Project Coordinator	LaRC/AMA

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		ISS Anomalies Trending Study	
		Page #: 8 of 48	

4.0 Executive Summary

The objective of this assessment was to utilize data mining and trending techniques to review the anomaly history of the International Space Station (ISS) and provide tools for discipline experts not involved with the ISS Program to search anomaly data to aid in identification of areas that may warrant further investigation. Previous NASA Engineering and Safety Center (NESC) data mining and trending assessments [ref. 1] performed analysis on data contained in individual anomaly recordkeeping systems (i.e., databases). However, ISS anomalies and nonconformances are documented in multiple databases. The assessment team prepared and integrated pertinent ISS nonconformance data from multiple sources and provided an enriched data set that was easier to navigate and use.

The data trending goals were to:


- Demonstrate the capability to trend ISS anomaly data from multiple data sets.
- Provide a means for discipline experts to gain deeper insight into ISS anomaly data.
- Provide fresh insight into ISS problem trends and significant anomalies, as able within the assessment timeline.
- Learn successful approaches to assist discipline experts in trending across multiple, merged data sets.

The timeframe for the assessment was approximately 1 year to accomplish these goals; however, the goals were not fully met. The preparation, integration, mining, and presentation of the ISS data took longer than expected, with little time left to perform in-depth analysis with the discipline experts. This report documents the activities completed to date and focuses on documenting the tasks of data preparation, integration, text mining, and visualization. Additional analysis of the ISS data is recommended and will continue outside this assessment.

The team completed extraction of pertinent data fields from the six nonconformance data sets and installed the merged data on a secure Microsoft® SharePoint® site, with security restrictions and controlled access. Colocating the nonconformance data from different reporting systems was an important first step in enabling trending analysis and data mining of the nonconformance records. The data sets included:

- Problem reporting and corrective action (PRACA) and items for investigation (IFI) data—both included in the ISS Problem Analysis Resolution Tool (PART)
- Government-furnished equipment (GFE) discrepancy reports (DRs) and GFE PRACA from the Quality Assurance Record Center (QARC)
- Mission Operations Directorate (MOD) Anomaly Reports (ARs)
- Software Change Requests (SCRs)
- Maintenance Analysis Data Set (MADS)

Given the different designs of these data sets, transformation of the data was necessary (i.e., storing it in proper format or structure to enable querying and analysis). In some cases, this

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 9 of 48	

was as simple as normalizing the names of like fields. In cases where fields were nonexistent for one or more of the data sets, this step was more complicated. The data sets all have free (unstructured) text fields (e.g., title, description) and prescribed (structured) fields (i.e., pull-down menus for trend code selection and many other types of selection). The IFIs and ARs, however, have few prescribed fields and do not include codes for types of failure modes, defects, or causes (e.g., requiring additional steps to improve the search).

Data- and text-mining approaches were used to enrich the data. These approaches convert information in text fields into indexing data or topics. The topics discussed in the text fields could then be used to search and filter the data sets, to find similar anomaly reports that might otherwise be missed. These topics could also be used to develop topic-based codes for failure modes, defects, or other commonly used codes in some of the data sets.


For data and text mining on individual or merged data sets, the goal was to provide discipline experts with better access to pertinent ISS anomaly data by converting topics from free-text fields into indexable data. Terms, concepts, and topics identified in text mining would be integrated into the merged data set to improve search for relevant reports.

- Statistical text and data mining would identify terms (often topics) in the text fields (e.g., in titles and problem descriptions) that were similar between correlated reports.
- Semantic text mining would identify concepts (topics) that occurred in text fields and use them to index reports in the data set. These topics would be taken from a large set of possible topics and would, therefore, be common across data sets. These topics also could be used to define standard proxies for trend codes such as failure mode codes. Trend codes are used slightly differently across some of the data sets and could be applied to all sets, including those where these codes have not been used.

Significant progress was made in the use of semantic text mining techniques to enrich the data and improve capabilities to search and filter reports. Semantic text mining uses a NASA tool, the Semantic Text Analysis Tool (STAT), which parses sentences in free text and then matches nouns, verbs, and modifiers with concepts (i.e., topics) that are represented in the NASA Aerospace Ontology. The ontology is a large hierarchical data structure that is designed to recognize multiple words and phrases used in free text in aerospace to denote thousands of types of entities, properties, actions, and problems. These concepts are equivalent to a common index for all reports in the merged data set. This text-mining approach was used and its accuracy was verified in a previous project [ref. 2] on analysis of DRs from QARC.

The results of the text analysis—a set of topics associated with each data record—were reported in formats that were integrated into the merged data set. A method was defined for using these topics to expand search to more relevant items, so that fewer of them would be missed in regular searches. This method has not yet been rigorously tested.

The set of topics associated with each data record was used to develop topic-based rules for proxy failure mode and defect code fields. This was a second use of the results of semantic text analysis. Establishing identical trend code fields across data sets aids standard search. It was

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 10 of 48	

also expected that proxy codes would help overcome the manual coding limitation to select only one code when multiple codes would be appropriate. GFE PRACA trend codes were chosen as the standard codes for all data sets. Several approaches for defining proxy codes were tried, including a statistical machine learning approach. Supporting extensions to STAT were developed, and additional software was developed for preliminary evaluation of the accuracy of the proxy codes during their definition. Proxy codes were delivered for the two PRACA sets: IFIs and MOD ARs.

During this development, cases of wholesale errors in some manual codes were discovered. It became clear that the manual codes should have been vetted. Given the low accuracy of some manual codes, the statistical machine learning approach, which was used to define rules for proxy defect codes, should be rejected until vetting of manual codes results in selection of accurate training sets. The extended nature of this work left little time for vetting and evaluation of the accuracy or helpfulness of the topics associated with each data record extracted by STAT.

Two types of tools were customized for searching, browsing, and visualizing the data set to provide multiple perspectives on the data, with the goal of supporting further independent analysis. Tableau[®], a search and data visualization tool for business analytics, was customized to provide data team members and discipline experts with interactive dashboards and multidimensional report browsers for exploring the merged data. It was demonstrated that Tableau[®] could be used to identify trends in nonconformances across the merged data set.

Flamenco, an open-source search and visualization tool for multidimensional search, was customized (Flamenco+) to use the hierarchical indexes provided by STAT and the Aerospace Ontology for the data sets. Flamenco+ was also adapted for evaluating codes and analyzing trends. Corresponding STAT adaptations were made to provide output to support use of Flamenco+ for evaluation of proxy codes. The NESC assessment team was not able to fully realize strategies for information retrieval based on concept tag indexing and multidimensional faceted search using Flamenco. Integrated use of Flamenco+ and Tableau[®] was not explored but is feasible and promising.


The SharePoint[®] site enables discipline experts to go to one location to access the data and to then search across the data sets simultaneously. Several topics were investigated in the enriched merged data set. They include nonconformances in Extravehicular Mobility Unit (EMU) water separator fan bearings and harmonic drive/peristaltic pumps. Initial limited analysis was performed for software, human factors, and electrical power systems. SAS[®] Text Miner was used for some analyses to capture topics mentioned in text fields and structured fields, to guide search. Slow integration of results of semantic text mining did not leave enough time to define and evaluate methods using this topic information. Lessons from exploration of these discipline areas have been documented to improve future trending and data mining.

Late in the project, a new use of the data by Safety and Mission Assurance (S&MA) personnel and other interested organizations was identified. The objective would be to relate anomaly record information from the ISS merged anomaly data set to potential risks and hazards defined in the ISS Hazard Analysis System. Given a hazard of concern or interest, historical anomalies

	<p align="center">NASA Engineering and Safety Center Technical Assessment Report</p>	<p>Document #: NESC-RP- 14-00950</p>	<p>Version: 1.0</p>
<p>Title: ISS Anomalies Trending Study</p>		<p>Page #: 11 of 48</p>	

that have occurred (that may have led to the occurrence of the hazard) and their risk ranking, perhaps by the number of related anomaly counts, could be compiled. These incidents may be reviewed, counted, and trended to raise awareness and to assess whether preventive actions would be prudent. The ability to search across several databases to identify relevant incidents is a key attribute to finding a more complete set of incidents for analysis. Further work is needed to define the use scenario and to evaluate the usefulness of the tools for this scenario.

This activity demonstrated use of the tool suite for deep investigations into technical issues related to focused problems. The team developed a tool suite framework (i.e., merged and enriched data, software, user interfaces, methodologies, processes, and practices) that can inform the potential expansion into other program/project data sets and support periodic updates of ISS problem-related data for ongoing interactive analyses by Technical Discipline Teams (TDTs).

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		12 of 48	

5.0 Assessment Plan

The objective of this assessment was to utilize data mining and trending techniques to review the anomaly history of the ISS and provide tools for discipline experts not involved with the ISS Program to search anomaly data to aid in identification of areas that may warrant further investigation. A challenge to investigating anomalies is that there are several problem reporting systems that hold data of interest, and the reporting systems do not have the same key data fields. The assessment team wanted to develop an approach to navigate through multiple problem reporting data sets simultaneously.

The assessment had four high-level goals:

- Demonstrate the capability to trend anomaly data utilizing multiple data sets.
- Provide a means for discipline experts to gain deeper insight into ISS anomaly data.
- Provide fresh insight into ISS problem trends and significant anomalies, as able within the assessment timeline.
- Learn successful approaches to assist discipline experts in trending across multiple, merged data sets.

To accomplish these goals, the assessment team established the following basic approach:


- Develop a method to capture integrated problem reporting data.
- Develop a capability to search for problem trends and effectively display meaningful trend data.
- Utilize semantic data mining to provide conceptual indexing and missing failure and defect codes.
- Establish a capability for discipline experts to search ISS data across multiple anomaly databases.
- Identify trends and significant issues from targeted reviews of software, electrical power, mechanisms, and human factors disciplines.
- Document the data mining and trending development effort to inform potential follow-on capability for cross-program/project trending.

6.0 Description of Data Sub-team Tasks

The NESC assessment team consisted of two subteams—the data subteam and the discipline expert subteam. Section 6.0 describes the data subteam’s effort.

6.1 Team Methodology

The data subteam prepared the nonconformance data for further analysis, delivered the initial analysis, and aided discipline experts with their investigations. The discipline expert subteam utilized the initial analysis and data/tools to further investigate for adverse trends or significant anomalies.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		13 of 48	

One of the major ambitions of the assessment was to create a tool suite that discipline experts could use to investigate anomaly history and perform data mining across multiple ISS anomaly databases. The NESC assessment team foresaw many potential uses, such as looking at data trends across multiple systems, supporting root cause investigations or unique technical assessments, or providing supporting data for looking at precursors to failures.

The NESC assessment team first established a concept of operations for discipline expert use of the search tool(s) (see Appendix A). The concept of operations shows how the merged data product can be used to serve discipline experts in researching issues concerning ISS anomalies. Four potential discipline expert use cases were identified to support development of the enhanced data-mining tool. These use cases are described in further detail in Appendix A.

- Scenario 1: Identify recurring anomalies and emergent risks.
- Scenario 2: Provide in-depth problem investigation in support of an NESC assessment.
- Scenario 3: Associate a potential issue or hazard to the historical operational anomalies or failures that could have led to the realization of the hazard.
- Scenario 4: Provide supporting data for precursor analysis.

Late in the assessment timeframe when the tool suite was maturing, the data subteam worked with discipline experts in the areas of software, human factors, electrical power, and mechanisms. The general interaction with discipline experts is illustrated in Figure 6.1-1. Once the initial set of anomaly data was extracted from the multiple source databases and merged, visualization tools were used to build views and dashboards to support the discipline expert analyses. This initial set of discipline experts provided feedback for tool enhancements.

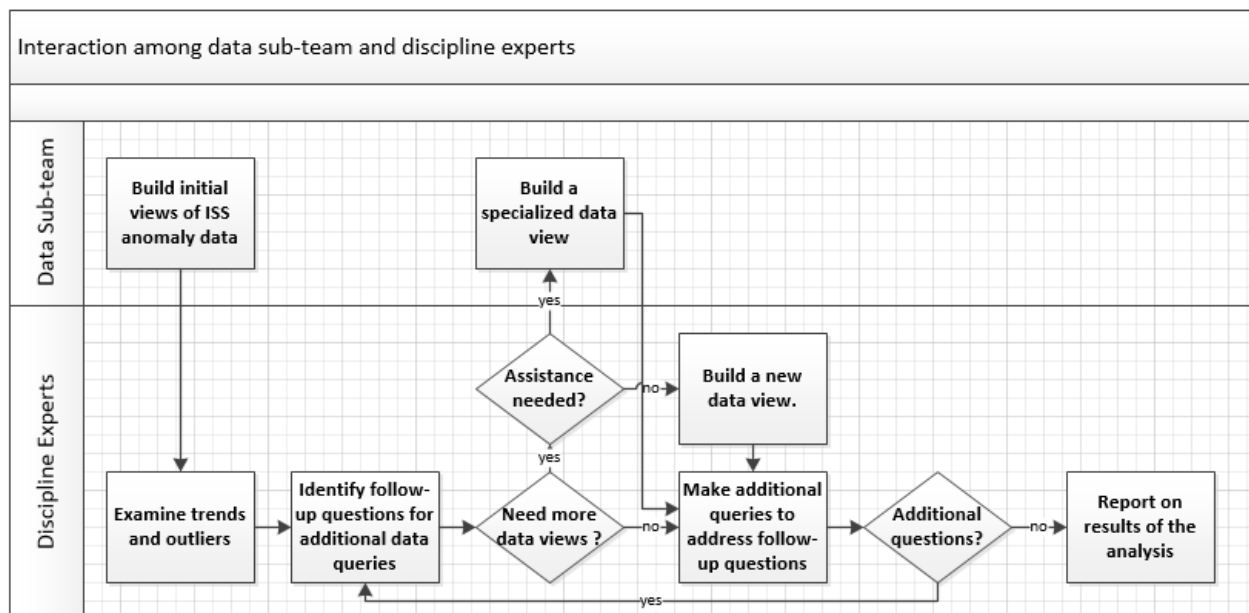



Figure 6.1-1. General Interaction with Discipline Experts to Support Analysis of ISS Anomalies

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		14 of 48	

6.2 Data Sources

6.2.1 Anomaly and Problem Reporting Data Sources

The data sources selected for this ISS assessment consisted of GFE DRs and GFE PRACA from the QARC, PRACA, and IFI data from the ISS PART, and MOD ARs. Each data source was selected by the NESC assessment team with the intent of providing data that would give insight into recurring or significant problems. The fields from these databases often did not overlap (i.e., freeform fields versus drop-down fields, handling of part numbers, serial numbers, etc.). This complicated merging of the data, as it limited which fields were selected for merger and drove effort to create new common fields, in some instances using the semantic mining techniques described later in this report. For instance, the MOD AR database generally had few fields compared with the other databases, and no defect codes or failure codes.

An additional complication was the manner in which each database handled anomaly reoccurrences. This is significant when trending counts of occurrences. MOD AR reoccurrences are typically added to an existing record with no indication that there is/is not a reoccurrence, or how many—the record must be opened and reviewed. Additionally, in some cases, records such as IFIs are upgraded to PART PRACAs and/or GFE PRACAs or DRs. This must be accounted for during analysis as well.


The nonconformance database record counts ranged from 3,992 to 220,006 records per data set. One of the main drivers for the differences observed in the counts across databases was the manner in which problems are recorded. Flight databases (i.e., PART PRACA/IFI and MOD AR) typically only generate a record against the offending part or problem, while the GFE data sets often delve deeper into a nonconformance and spawn separate records for the subcomponents and/or all serial numbers of an offending part and/or its components.

6.2.2 Additional Data Sources

Additional data sources were made available to further support anomaly investigation. These included the SCR data and the MADS. The SCR data provided deeper insight into flight problems that were transferred there for further troubleshooting or, in some cases, design changes. The MADS data were used to gain insight into the hardware that was or had been on orbit (see Table 6.2-1). Fields from both were added to Tableau® to provide cross-referencing while performing search and visualization.

Table 6.2-1. Ancillary Data Sources

Ancillary Data Sources	
Data Sources	Record Count
SCR	40,361
MADS	1,921

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		15 of 48	

6.3 Data Extract, Transform, and Load (ETL)

A goal of this assessment was to perform data mining across multiple data sources. To establish this capability, data had to be extracted from each data source. The data were transformed into a common set of fields and loaded into a single database, enabling data mining and trending. This multistep process is referred to as ETL and is shown in Figure 6.3-1.

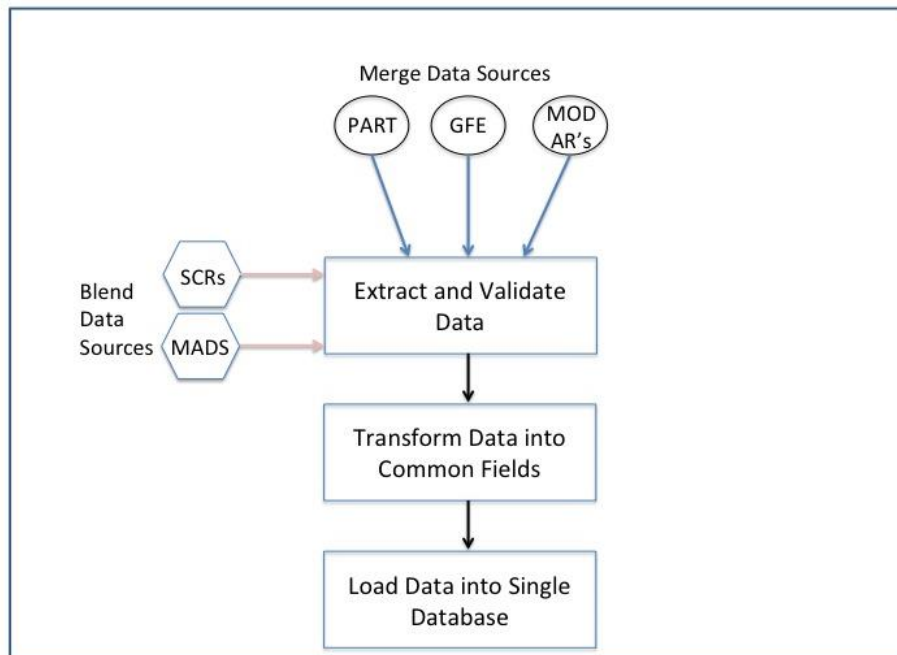


Figure 6.3-1. ISS Data Sets Extraction, Transformation, and Load

6.3.1 Extract

Data extraction is the act of retrieving data from your desired data sources for further processing and subsequent storage. Extracting data from ISS data sources had challenges because of differing formats, security access, and understanding how the various fields were used (e.g., fields with the same name may have different content, and fields with the same content may have different field names).

Nonconformance records from the GFE DR, PART PRACA, and PART IFI databases were extracted using their web interfaces by running a single report that was output in Excel[®] format. Accessing MOD AR data was more challenging because it was accomplished by running reports from the database web interface for each ISS increment and then exporting the individual nonconformances in the increment report to an Excel[®] file. Each Excel[®] file was then combined into a single file. Access and extraction of data from the data sources required contacting the data owners, requesting access to the data, and meeting the owner's security requirements.

The data extracted from the data sources were static, so the data were current only from the day the data were retrieved. Table 6.3.1-1 shows the data extraction date for each record system.


	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		16 of 48	

Table 6.3.1-1. Data Extraction Date for Each Record System

Data Set	Extraction Date
GFE DR	September 24, 2014
GFE PRACA	June 24, 2014
PART PRACA/IFI	January 7, 2015
MOD AR	January 7, 2015
MADs	June 30, 2014
SCRs	July 31, 2014

The extraction of GFE PRACA was performed using a standalone Microsoft® 2008 Server and by building a Microsoft® SQL 8 database. The data set had Shuttle data and crossover data (i.e., both ISS and Shuttle), so a Microsoft® query was built to extract only ISS data. The queries were improved as the NESC assessment team vetted the data. For example, some adjustments were needed when it was noticed that not all of the extravehicular activity (EVA) data were retrieved in the initial queries. This was found during early analysis and corrected.

SAS® Enterprise Guide was used to review and set up the large data sets that combined visualization and search in Tableau®. Tableau® visualization was used for early data quality control. Data discrepancies were easier to find using visualization.

Data owners were instrumental in providing road maps to the data and providing the documentation required to help the team make decisions on which fields to use. They provided data code manuals, reports, supporting documentation, and data dictionaries.

The initial extraction included 353 fields from five different problem reporting data sources. After review by the data subteam, the number of fields was reduced to 209 fields. The data subteam further consolidated those into 36 fields. This review identified fields required to combine five different ISS problem reporting data sources into one source. Many of the discarded fields were system-generated fields that controlled the document status or the date and time transaction. Additionally, many of the excluded fields were specific to processing data within that data source, as in a document workflow.

There are two distinct field types: structured fields and unstructured fields. Structured fields have predetermined options available for selection (e.g., codes and code descriptions). Usually, these are in dropdown menus that a user has to select. Unstructured data accept freeform data, with little or no organization. For example, a field entitled “problem description” typically allows freeform entry of a prescribed amount of characters. These free text fields caused challenges for data mining, due to spelling errors, acronyms, special characters, and other text irregularities. There were four unstructured fields used in the combined data set: Problem Title, Problem Description, Detected During, and Part Description, as shown in Table 6.3.1-2. This table also lists the structured fields used to separate problem reporting documentation into categories that could be searched for trending and analysis.


	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		ISS Anomalies Trending Study	
		Page #: 17 of 48	

Table 6.3.1-2. Structured and Unstructured Fields Used in Merged Data Set

Reporting Codes (Structured Fields)		
• Program	• Subsystem	• Defect
• Project	• Flight Element	• Failure Mode
• Cause	• System	• Prevailing Condition
• Disposition	• Test Operation	• Recurrence Control
Unstructured Fields		
• Problem Title	• Problem Description	• Detected During
• Part Description		

Even though many fields were not used in the merged data set, links were provided (in Tableau®) to the original data source and added to each record to provide for a more in-depth analysis of individual records, if necessary. Each complete record could be viewed by following links to the original data source web site.

6.3.1.1 Extraction of Additional Data Sources Fields

Two additional data sources were used to further support nonconformance data analysis, the SCR and the MADS (see Table 6.3.1-3) data sets. SCRs document software updates (which are sometimes kicked off via nonconformances) and MADS are used for capturing hardware maintenance activities. The MADS and SCR data sources were then blended with the related problem reports. Blending did not result in adding fields to the merged data set, but supported looking up related details. For example, a problem report may refer to a part number that could then be examined further by searching the MADS data.

Table 6.3.1-3. SCR and MADS Data Fields

SCR Data Fields	MADS Data Fields
<ul style="list-style-type: none"> • Reason for Change • Subsystem • Test Environment • ISS SCR Number • Status • Provider • Originator Stage • CSCI • Created Date • Title • Board 	<ul style="list-style-type: none"> • Part Number • Location • Flight Activated • Unique ID • Old Part Number • Part Name • Hardware Criticality • Flight Manifested • Type Name • System • Function • EVA or IVA Overhead Time • Type of Part

6.3.2 Transform

6.3.2.1 Data Source Fields Transformed

The review of the five data sources consolidated 353 fields into 36 transformed fields.

These fields were chosen based on the relevance of the data for trending and subsequent insight into trends and significant problems. Where there were different field names with the same data types, those field names were transformed as shown in the example in Figure 6.3.2-1.

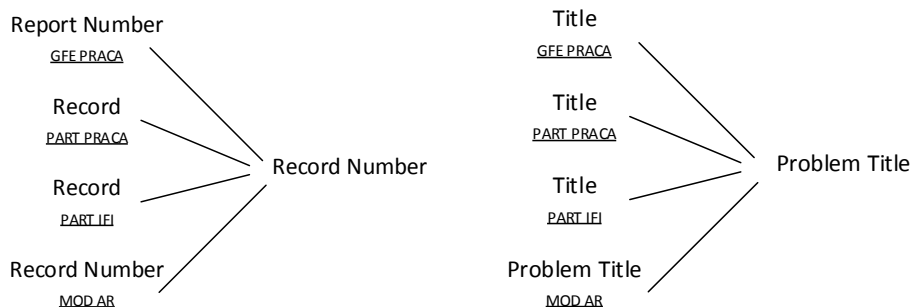



Figure 6.3.2-1. Transformed Fields Examples

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		19 of 48	

A new field, Database Name, was added to help facilitate record integrity where records from two data sources had the same data identifiers but had no relationship, as in Record Numbers with PART IFI and MOD AR. Figure 6.3.2-2 shows the methodology that was used to maintain record integrity when combining those data.

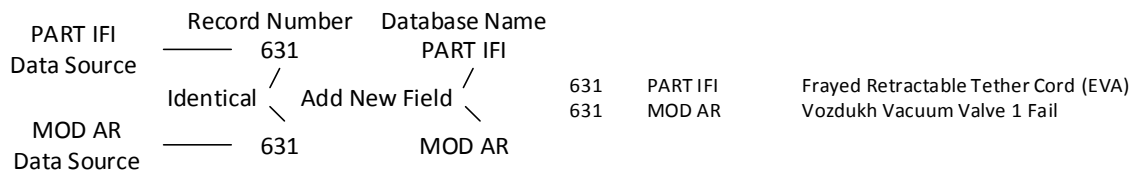


Figure 6.3.2-2. Field Addition for Record Integrity Example

The full list of transformed fields is shown in Table 6.3.2-1, including three added fields (Sub Ontologies, CTags (concept tags), and CTag Count), which are explained in Section 6.4.3.1.


Table 6.3.2-1. Transformed Fields

- Record Number
- Originator
- Status
- Program Code
- Detected Date
- Detected During
- Disposition Code
- Manufacturer
- Prevailing Condition Code
- Recurrence Control Code
- Test Operation Code
- Flight Element Code
- Sub Ontologies
- System Code
- Site Location
- Hardware Type
- Flight
- Defect Code
- Defect Description
- Failure Mode Code
- Failure Mode Description
- Responsible Org
- Activity
- Hardware Ownership
- Problem Description
- CTags
- Project Code
- Problem Title
- Cause Code
- Cause Description
- Like HW On Orbit
- Part Number
- Part Description
- Serial Number Lot
- Database Name
- Related Document
- Subsystem Code
- Subsystem Description
- CTag Count

6.3.3 Load

6.3.3.1 SAS® Data Load for Data Visualization

The completion of the transformed and combined fields brought the NESC assessment team to the next phase, which involved converting the data into a file that the Tableau® Desktop software

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		20 of 48	

could import. Because a standalone version of Tableau® was utilized, a Microsoft® Excel® file was needed to make the data portable for the Tableau® Reader. This allowed the Tableau® file to be downloaded to any desktop or laptop to review the entire transformed database. The software used for the data conversion to Microsoft® Excel® was SAS® Enterprise Guide (EG). This was the same software that was used for the transformation phase of ETL. Workflows were set up using EG so that new data could be added or modified as needed. SAS® EG was used during data refresh to add descriptions to field coding (i.e., cause, defect, failure, subsystem descriptions), when data were updated, and during vetting.

6.4 Tool Suite

There is no “perfect” tool for identifying trends or significant anomalies. Overlapping techniques are necessary to improve results. Overlapping techniques are useful when working with the nonconformance data sets to rule out irrelevant reports, remove duplicates, corroborate relevant reports, and identify reports that were expected but not found. The resulting data set can then be counted and presented in time-related trends.

The data subteam’s approach was to utilize a merged data set and apply data-mining tools and techniques to enhance the ability to identify trends and significant anomalies by applying a suite of capabilities. The methods used to explore nonconformances included (1) search, (2) improved search by way of adding concepts to anomaly reports (concept tags), and (3) adding failure mode and defect code fields using “proxy codes” to nonconformance data sets that did not have them (i.e., MOD AR and PART IFI). Several tools were used for searching and visualizing the data: Tableau®, Flamenco, and SAS®. Statistical text mining using SAS® identified correlated documents, based on terms they have in common, to find reports that may be missed using full text search. SAS® was also used to update the Aerospace Ontology, which was used in conjunction with the STAT to develop the concept tags and proxy codes. Flamenco, enhanced to become Flamenco+, was used for its strength as an open-source faceted search and visualization tool. Tableau® was used for its strength as an intuitive, state-of-the-art data visualization tool.

6.4.1 Search

Full text search is a common information retrieval method when key information for selecting reports is in text fields. Common search strategies are iterative and interactive to give the user an opportunity to improve the search query until the sought-for item is found. Using this strategy with ISS anomaly data sets is useful, yet insufficient by itself; it is relatively easy to judge whether a report is relevant, but finding the right reports is difficult.

The most common reasons for failing to retrieve reports with search are word variations, which include synonyms, multiple spellings and misspellings, abbreviations, acronyms, and other shortened forms. An automatic query reformulation or search expansion strategy could help overcome the problem of word variations if these variations can be collected from the text in the data set. STAT and SAS® also provided spelling correction and stemming to base forms (e.g., “closing” changed to “close”). This collection strategy was used early in the development of the merged data sets prior to the utilization of data-mining tools. Simply using search on merged

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		21 of 48	

data sets added value compared with searching nonconformance databases separately. The same results could have been achieved by combining these search results using the latter approach; however, that approach would have been considerably more cumbersome and time consuming.

6.4.2 Data Mining to Enhance Search

Figure 6.4.2-1 shows the activities the NESC data subteam performed to enhance the nonconformance reports by adding proxy defect and failure codes and “concept tags.” It shows the stages of transformation from the original data sources to the final merged data views, including enhanced search, visualized using Tableau® and Flamenco+. Some data sources (e.g., GFE PRACA and PART PRACA) had problem reporting codes (e.g., failure mode codes and defect codes) that could be selected from pull-down lists. PART IFI and MOD AR data sources did not have failure mode or defect codes. These data fields were created for PART IFI and MOD ARs using proxy codes, which enable searching with these codes simultaneously across all data sets.

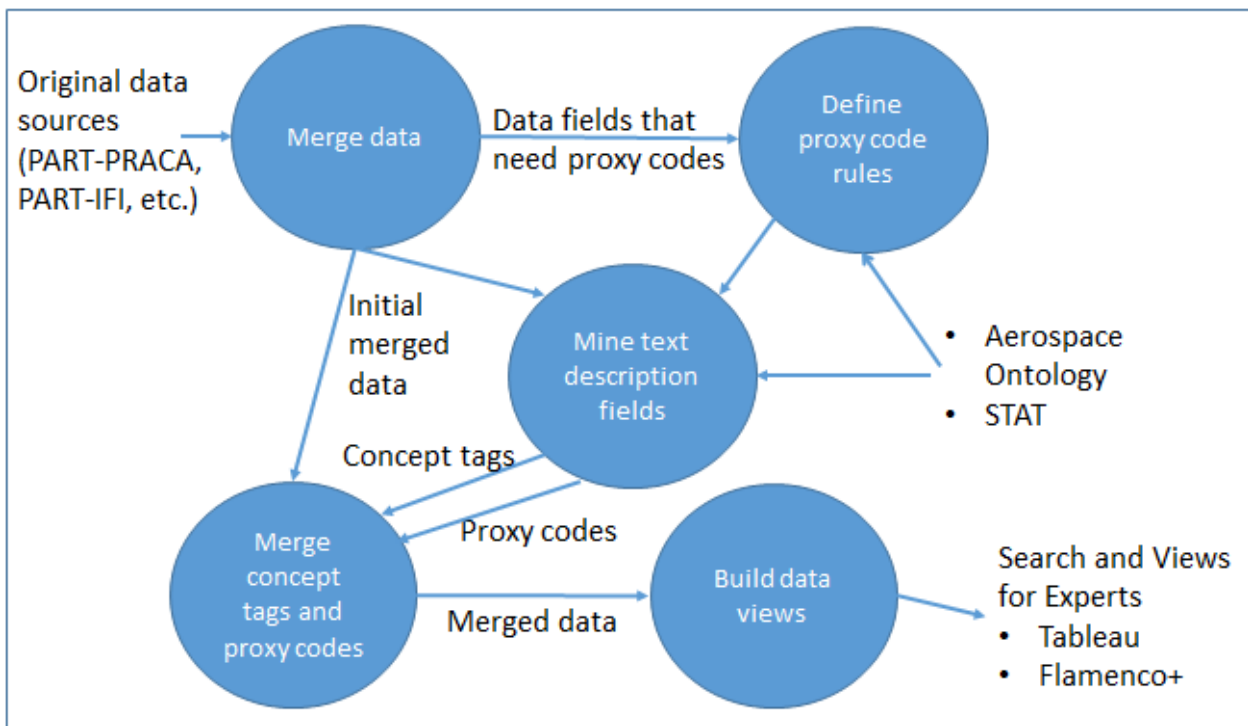



Figure 6.4.2-1. View of NESC Data Subteam Activities

The Aerospace Ontology and STAT were used to develop concept (topic) tags. The concept tags were intended to enrich each anomaly report by adding relevant concepts or topics to individual nonconformance reports, improving the ability to group nonconformances when searching. The concept tags are assigned based on analysis of the text from unstructured (i.e., free-text) fields: the Problem Title and Problem Description fields. Likewise, rules for assigning proxy codes were developed using the concept tags. The concept tags and proxy codes were added to the merged data set and used directly in the data views in Flamenco and Tableau®.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		22 of 48	

6.4.2.1 Aerospace Ontology, Concept Tagging, and Proxy Codes

6.4.2.1.1 *Concept Tagging*


Semantic text mining with STAT and the Aerospace Ontology identifies and tags reports with the concept-topics that are mentioned in the Problem Title and Problem Description text fields of each report. The goal of concept tagging is to provide discipline experts with better access to pertinent ISS anomaly data by extracting concept-topics from free-text fields so they can be used to index, search, and filter reports in the merged data set.

The Aerospace Ontology concept-topics are equivalent to a common index for all reports in the merged data set. Each concept-topic in the Aerospace Ontology is associated with a list of terms (words and phrases) and variants that represent that concept so that it can be matched with nonconformance free-text fields. These indexing concepts are robust to many variations in the way topics are expressed in text. The Aerospace Ontology contains thousands of indexing concepts and tens of thousands of terms, which have been developed over years of effort, most recently with GFE nonconformance records (i.e., DRs). The structure of concepts in the Aerospace Ontology is hierarchical and is organized at the top level into sub-ontologies for types of properties, objects, actions, and problems in the aerospace domain.

Prior to using the Aerospace Ontology to develop concept tags, concepts and terms were added to the Aerospace Ontology for the ISS nonconformance domain. Methods were developed and used successfully to semiautomatically identify new terms and variants (from the merged data set) to add to the ontology. Lexical analysis of the vocabulary in the merged data set, described in Appendix B, identified about 170,000 words and phrases to consider. A matching and frequency-ranking method, described in Appendix C, identified a set of less than 350 new terms that had priority to be added to the Aerospace Ontology. The version of the Aerospace Ontology that was used for indexing by text mining included these terms, as well as others identified during preliminary vetting of proxy codes. A spreadsheet-based procedure for adding new concepts and terms to the Aerospace Ontology is described in Appendix D.

Semantic text mining with STAT identifies and tags reports with the Aerospace Ontology concept-topics. STAT performs spelling correction and parses the content of the text fields to derive syntactic phrase structures with nouns, verbs, and associated modifiers. STAT then finds semantic (meaning) matches to concept-topics, based on lists of words or phrases that are associated with each concept. These matches are used to identify types of problems, objects, and properties in the text. One or more problem, object, or property concept-topics can tag each text field in each anomaly report. The results of the text analysis—a set of concept-topics associated with text fields in each data record—are output in table formats that were integrated into the merged data set.

STAT matches and indexes the Aerospace Ontology words and phrases by using the stemmed base forms of discrepancy words. This simplifies the matching to search for BAD or NO nouns or verbs. For example, “inadvertently closed” would be simplified to “BAD close.” Near matches such as “incompletely closed” would also be a type of “BAD close.” The phrase

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 23 of 48	

structures of each sentence guide the association of words within the phrases, such as in a case where there are intervening words between “inadvertent” and “close.” This simplifying strategy improves performance but can merge types of bad properties that need to be distinct. The resulting concept tag distinguishes the type of operation/function better than the type of problem property. This weakness can be remedied in the future by text analysis changes or by using a negative property dimension in faceted search.

In practice, STAT does not tag *all* of the ontology concepts in the text. Some concepts are too general. Others are unlikely to be of interest to the analyst. The configuration specifies a set of intermediate-level concepts (the “start-with-nodes”). STAT tags these concepts and the concepts below them.


This text-mining approach was used and its accuracy verified in a previous project on analysis and text mining of DRs from QARC. For more detail, see reference 2.

6.4.2.1.2 Development of Proxy Codes

STAT and the Aerospace Ontology were also used to develop proxy codes. The set of Aerospace Ontology concept-topic tags associated with each data record was used to develop concept-based rules for proxy failure mode and defect code fields. The proxy codes provide substitute failure mode and defect codes in the MOD AR and PART IFI data sets, where manual problem reporting codes are not built in. Analysts can search for similar records using these codes simultaneously across data sets. GFE PRACA trend codes were chosen as the standard codes for all data sets. These were chosen rather than the PART PRACA trend codes because there were fewer possible GFE PRACA trend codes, and they were more recent versions of the failure mode and defect codes.

Synthetic codes would serve as proxies for the missing manual codes. A plausible approach to generating these proxy codes was to define classification rules, using “and/or/not” logic based on the presence or absence of specified concept tags. The concept tags could be used as inputs to the rules. These rules would classify each trend code in a record into one or more proxy code values. Allowing more than one proxy code value per trend code could be useful in overcoming the problem of constraining manual codes to only one code per field when two or more codes would have improved the search for trends.

Several approaches for defining proxy codes were tried, including a statistical machine learning approach for defect codes. Proxy codes were assigned based on concept tags from the text in the Problem Title or the Problem Description field. Preliminary estimates of proxy code recall (i.e., the proportion of records with a particular GFE PRACA manual code found with the corresponding proxy code assigned) were about 30 percent. This rate is similar to the estimated likely manual recall (if assessed by trained judges, allowing multiple code values). The highest precision (i.e., proportion of assigned codes that matched a particular GFE PRACA manual code) for defect proxy codes was 0.27 (Mean = 0.10) and for failure mode proxy codes was 0.84 (Mean = 0.16). To improve precision, records with more than five proxy code assignments were reduced to the five codes with the highest precision in the initial measure of proxy code

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 24 of 48	

precision. More detailed descriptions of methods for proxy code development and refinement are provided in Appendix A (Section A.3.3) and Appendix F.

The inherent limitations and inaccuracy of the manual trend codes made it difficult to develop accurate classification rules for the proxy codes. These limitations and possible remedies are discussed further in Appendix F (Section F.5).

6.4.2.2 Search Using Concept Tags


Rather than building proxy codes from concept tags, the more promising approach is to use concept tags directly to search and browse. The concept tags were concatenated into a single string and made into a concept tag data field in Tableau® so that Tableau® users could search with Aerospace Ontology concepts to find anomaly records with similar attributes. The Tableau® visualization tool (see Section 6.4.3.1) can present multiple dimensions but does not currently support hierarchical faceted search as seen in Flamenco+. Tableau® performance problems have prevented full use of this search strategy.

The concept tags that are extracted from text fields are also a source of dimensions for faceted search. Faceted search combines keyword search with browsing in a multidimensional (i.e., “multifaceted”) hierarchical space. Analysts can begin with a classic keyword search and then scan the list of results while inspecting a display of related dimensions that provides insights into the content and its organization. The purpose of faceted search is to help the analyst determine quickly what types of attributes or dimensions are available and the counts of reports that contain concepts in those dimensions (see Section 6.4.3.2). The dimensions partition the items in multiple ways so that each anomaly report can be a member of several different groups of related reports. Combinations of dimensions and search within groups can filter sets of related reports into more specific subsets that target the trends of interest to the analyst.

Concept tags were implemented near the end of the assessment and not utilized enough to fully test their efficacy. They are incorporated in Flamenco+ and Tableau and at the very least will improve the ability to perform deep dives on particular topics where a search needs to be as comprehensive as possible. Given that, it is also expected that the concept tags will help improve the overall speed and accuracy of performing searches in general.

6.4.2.3 Search Using Proxy Codes

The purpose of proxy codes is to approximate what would have been assigned by a manual entry in the data sources where manual problem reporting codes were not used (i.e., MOD AR and PART IFI). If all merged data records included these codes, similar records from all sources could be retrieved in a similar manner. Data fields needing proxy codes were identified (i.e., failure mode codes and defect codes). STAT and the Aerospace Ontology were used to match concept (topic) tags with text in the title and description fields. The project used the four data sets to develop and evaluate proxy codes. The inherent limitations of the original codings (see Section 6.4.2.1) limit their usefulness for data discovery. This was found to be true. The limitations primarily stemmed from the inadequacy of the existing manual condition codes found in the existing data sets (i.e., GFE and PART PRACA). Significant manual coding errors were

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 25 of 48	

found during the process of developing the proxy code rules. Even though the team attempted to overcome this, preliminary estimates of proxy code recall (i.e., the proportion of records with a particular GFE PRACA manual code found with the corresponding proxy code assigned) were about 30 percent.

A more detailed description of methods for proxy code development and refinement is presented in Appendix A (Section A.3.3) and Appendix F.

6.4.2.4 Statistical Text Mining Using SAS

The purpose of the SAS[®] analysis text-mining phase was to identify reports for specific suspected problem areas, disciplines, or subsystems that could not be found easily with keyword search. Discipline experts specified lists of terms and noun groups that defined areas of focus. Statistical text mining was used to identify correlated documents, based on terms and noun groups they had in common. Each group of correlated documents represents a latent topic, which is defined by the common terms. Thus, new terms or noun groups could be identified to add to search expressions, if desired. The analysis was used to determine significant observations or trends that needed further investigation. For detailed information on SAS[®] analysis, see Appendix G. This approach proved useful for identifying potential areas of interest based by grouping similar anomaly topics. Since the methodology used in SAS is statistical based on word frequency, many of the clusters of anomalies identified turned out to be uninteresting. Consequently, wading through identified clusters is time-consuming and often uninformative.

6.4.2.4.1 SAS[®]


SAS[®] advanced analytics software packages used in this assessment were SAS[®] Enterprise Miner, SAS[®] Text Miner, and SAS[®] Enterprise Guide. SAS[®] Enterprise Guide was used to combine and transform the five different data sources, as discussed in Section 6.3.3.1. These software products were also used during the analysis phase for lexical analysis and to perform text mining to identify topics that could be used to find relevant reports that might be missed in search.

6.4.3 Data Visualization

A key enabler of data trend analysis is to have an effective tool for users to query the data and to visualize the output. This assessment used two complementary data query and visualization tools: Tableau[®] and Flamenco+.

6.4.3.1 Tableau[®]

Tableau[®], a commercial off-the-shelf (COTS) tool, was used for its strength as an intuitive, state-of-the-art data visualization tool. Tableau[®] Desktop is a multi-platform software program procured to assist the NESC assessment data team developers implement data visualization. Tableau[®] Reader is freeware used to connect to the data sources (i.e., merged data sets), which were built using Tableau[®] Desktop. Tableau[®] Reader was used by the discipline experts and the data subteams. Tableau[®] Desktop provided the capability for querying, calculating, code

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 26 of 48	

generating, and graph building for the construction of data visualization dashboards, saved as Tableau® files. The Tableau® Reader is used to interact with these files, providing a viewing capability, querying, filtering, sorting, exporting, and printing. This facilitates the interactive visualization of the files produced by the Tableau® Desktop component.

The data visualization dashboards were designed and developed to provide quick access to the multidimensional aspects of the information contained in the nonconformance reports. The dashboard shown in Figure 6.4.3-1 depicts six zones of interest: one primary query zone and five display zones. Following the numbering on the figure, Zone 1 is a text entry area used to query the combined data sets. Zone 2 summarizes record counts over a trending timeframe (a record is a single nonconformance record, e.g., PART record 9202) showing occurrences detected per year and total records per database. Zone 3 is the records table, which includes title, description, and link to the original record database. Zone 4 contains various other counts, such as a count by part number and a count by cause codes. Zone 5 shows records related to the currently selected record, as well as an ability to filter records by cause, defect, or failure mode. Zone 6 contains the concept tags and includes a text entry area to filter the concept tags down to those tags containing the entered text, and additionally filters all other zones on the dashboard simultaneously. See Appendix E for more details on the zones in the figure and for further explanation of the use of Tableau®. The user manual in Section E.1 of Appendix E details additional Tableau® dashboard functionality.

The merged data set provides the ability to trend across a broader data set, and Tableau® makes it straightforward and intuitive to view the data. There is overlap between the data sets that often skews the counts, however. This adds burden to the user to manually remove duplicates once identified. For instance, at times a nonconformance identified in the MOD AR data set results in a nonconformance in the PART IFI data set, which may then end up in one of the GFE data sets.

Tableau® has proven useful for search and discovery. It is also valuable for exporting data/information to other tools such as Microsoft® Excel®, where additional cleanup, reduction, or formatting can be performed.



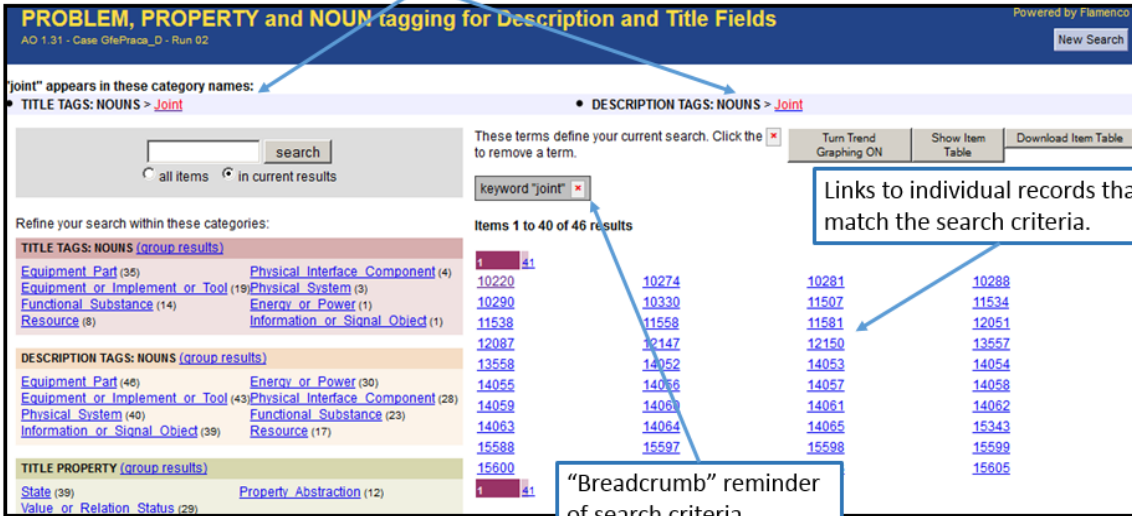
Figure 6.4.3-1. Data Visualization Dashboard

6.4.3.2 Flamenco

Flamenco, enhanced to become Flamenco+, was used for its strength as an open-source faceted search and visualization tool. “The faceted search model leverages metadata fields and values to provide users with visible options for clarifying and refining queries. It features an integrated, incremental search and browse experience that lets users begin with a classic keyword search and then scan a list of results (or do additional search). It also serves up a custom map (usually to the left of results) that provides insights into the content and its organization and offers a variety of useful next steps. That’s where faceted navigation proves its power. In keeping with the principles of progressive disclosure and incremental construction, users can formulate the equivalent of a sophisticated Boolean query by taking a series of small, simple steps. Faceted navigation addresses the universal need to narrow a search. Consequently, this pattern has become nearly ubiquitous in e-commerce, given the availability of structured metadata and the clear business value of improving product find-ability” [ref. 3].

The Flamenco+ faceted search environment is customized to show concept facets in the area to the left of the results of a faceted search. The search for “joint” in Figure 6.4.3-2 identifies 46 reports where the word “joint” appears in the text, and two “joint” (as a noun) concept-topics, one from the Title field and one from the Problem Description field in GFE PRACA records. Clicking on these links will lead directly to the set of reports that are tagged with this Joint concept tag. This will identify reports where the word “joint” or one of its 19 variants appears in the text. The variants include such terms as “SARJ,” “slip joint,” “join,” and “coupling.”

Shows “Joint” appears in both Title tags and description tags – Nouns part of Ontology




The screenshot displays the Flamenco+ search interface. At the top, it shows the search title "PROBLEM, PROPERTY and NOUN tagging for Description and Title Fields" and the search term "joint". Below the search bar, there are faceted navigation options for "TITLE TAGS: NOUNS > Joint" and "DESCRIPTION TAGS: NOUNS > Joint". A search box with "keyword 'joint'" is visible. The main results area shows "Items 1 to 40 of 46 results" as a list of report numbers. On the left, there are faceted navigation panels for "TITLE TAGS: NOUNS (group results)", "DESCRIPTION TAGS: NOUNS (group results)", and "TITLE PROPERTY (group results)".

Annotations on the screenshot include:

- A box pointing to the search results: "Links to individual records that match the search criteria."
- A box pointing to the search criteria: "'Breadcrumb' reminder of search criteria."

Figure 6.4.3-2. Results of a Flamenco+ Keyword Search for “Joint”

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		ISS Anomalies Trending Study	
		Page #: 29 of 48	

The facets on the left of the figure provide a custom map of the concepts associated with the 46 reports retrieved by keyword search, in order of frequency. Under each facet can be seen other classifications that are associated with “joint”; in the “Title Tags: Nouns” facet, the user can see that “joint” cross-cuts many concepts (e.g., “equipment part,” “physical interface component,” “energy or power”). From here, the user may want to select “equipment part” under “Title tags,” the most frequent category, to refine the query. Alternatively, the user can choose to perform another search (e.g., for “locking”) to refine the 46 results further. This scenario is discussed in detail in Appendix E.

The facets were designed to navigate by selected concept dimensions (from the Aerospace Ontology) or by type of code (failure mode and defect code). In this design, there are six ways to browse or filter based on the type of concept tags in various text fields (i.e., title or description field × object/noun, property, or problem). Six more facets support vetting of proxy codes (i.e., title or description field × manual or proxy × failure mode or defect code). These facets are illustrated in Appendix E, Figure E-14. Many other facet designs are possible for the ISS anomaly data set.

Due to the late incorporation of proxy tags, Flamenco+ was not utilized for search during the assessment but is available for use going forward. Flamenco should be optimized for search using concept tags.

6.5 Products Used, Purchased, and/or Developed

6.5.1 Data Sets and Data Set Documentation


6.5.1.1 ISS Anomaly Data Sets

The final ISS anomaly data set included:

- Combined anomaly records, as depicted in Figure 6.3-1.
 - Including Failure and Defect proxy codes that were added to records originally without these codes.
 - Including concept tags.

6.5.1.2 Aerospace Ontology Data

STAT semantic annotation or “tagging” relates parts of the text to concepts in the Aerospace Ontology, a lexicalized ontology. In a lexicalized ontology, each concept is associated with a list of words or phrases that are possible text representations of the concept. The Aerospace Ontology is implemented in Protégé. The final Aerospace Ontology version that supported STAT processing and delivery of concept tags and proxy codes is AO1.31 (.owl) and Version 1.31 Aerospace Ontology (.xml). Versions of the Aerospace Ontology that were developed during this project (in both .owl and .xml formats), in addition to V1.31, are only available upon request. Please contact the NESC at NESC@nasa.gov.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		30 of 48	

6.5.1.3 STAT Text Mining Result Data

6.5.1.3.1 *Concept-topic Tags*

These tags are available upon request. Please contact the NESC at NESC@nasa.gov.

6.5.1.3.2 *Proxy Codes*

These codes are available upon request. Please contact the NESC at NESC@nasa.gov.

6.5.2 Software and Software Reference Documentation

The following items were purchased, or downloaded as open source: Protégé, SAS®, Tableau®, and Flamenco.

6.5.2.1 Data and Text Mining Software

Protégé: Protégé is open-source software for editing ontologies and building intelligent systems. The software (V4.3) can be downloaded at <http://protege.stanford.edu/products.php#desktop-protege>.

Plugins for spreadsheet-based updating, XML output, and acronym checking are available upon request. Please contact the NESC at NESC@nasa.gov.

SAS® Software Tools


The following SAS® software tools were purchases under this assessment:

SAS® Analytics Pro v9.4: SAS® Analytics Pro 9.4 is the foundation of Base SAS® that houses the SAS® (data management facility, programing language, data analysis, and reporting) database and programs (Enterprise Guide, Enterprise Miner, and Text Miner).

SAS® EG v6.1: This point-and-click interface generates code to manipulate data or perform analysis automatically and does not require SAS® programming experience to use. SAS® EG provided the functionality that allowed us to perform ETL functions of the data into a homogeneous data structure. Because the data resided in many different heterogeneous databases and formats, SAS® EG helped to facilitate the extraction of data from many Microsoft® Excel® files and transform these data into a more homogeneous data structure, and to load export data into Tableau® readable files for Tableau® visualizations. SAS® EG also provided the path to update the files from the data sources early in the performed workflow process by putting into place several parameters. This expedited the entire process.

SAS® Enterprise Miner v13.1: SAS® Enterprise Miner streamlines the data-mining process to create predictive and descriptive models based on analysis of vast amounts of data. Enterprise Miner and Text Miner provided capabilities to explore and discover information found in the many textual data fields. This enabled the consolidation of the information into concepts and clusters.

SAS® Text Miner v13.1: SAS® Text Miner tools enable information extraction from a collection of text documents to uncover the themes and concepts.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		31 of 48	

STAT Semantic Text Analysis Tool

STAT is a syntactic parser and semantic interpreter and tagger implemented in Perl and Lisp that uses flat files as input and output. A STAT tar file is available upon request. Please contact the NESC at NESC@nasa.gov.

Ontology Updating Software

The Guide for updating the Aerospace Ontology based on lexical corpus analysis is contained in Appendix C of this report.

The Python software for performing this updating is available upon request. Please contact the NESC at NESC@nasa.gov.

Proxy Code Development and Evaluation Software

Python scripts were developed to export Flamenco+ concept tags, generate proxy code rules, and evaluate their precision and recall. This software is available upon request. Please contact the NESC at NESC@nasa.gov.

6.5.2.2 User interface and Visualization Software

Tableau® Software: Tableau® Desktop is a multi-platform, COTS software program procured to assist the NESC assessment team developers in implementing data visualization.

Tableau® Reader: Tableau® Reader is freeware used to connect to data sources (merged data sets) that were built using Tableau® Desktop.

Flamenco+: Flamenco is a search interface framework implemented in Python using a MySQL database and is available at <http://flamenco.berkeley.edu/index.html>.

Flamenco+ was developed to enhance the user interface and output capabilities for searching and browsing problem reports and other NASA short documents. A Flamenco+ tar file is available upon request. Please contact the NESC at NESC@nasa.gov.

6.5.3 Guides and Training Products

6.5.3.1 User Guides

6.5.3.1.1 Flamenco+ User Guide and Tutorial


A Flamenco+ User Guide and Tutorial is available upon request. Please contact the NESC at NESC@nasa.gov.

6.5.3.1.2 Tableau® Dashboard Tutorial

The Tableau® tutorial is contained in Appendix E, Section E.1, of this report.

6.5.3.1.3 Data Mining Site Users Guide

The “ISS Data Mining Site Construction Guide” is contained in Appendix H of this report.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		32 of 48	

6.5.3.2 Developer Guides

6.5.3.2.1 *Ontology Customization Guide*

The Ontology Customization Guide is contained in Appendix D of this report.

Previously developed user guides for inspecting and updating the Aerospace Ontology are available upon request. Please contact the NESC at NESC@nasa.gov.

6.5.3.2.2 *STAT Analysis Tutorial and User Guide*

A STAT Analysis Tutorial and User Guide are available upon request. Please contact the NESC at NESC@nasa.gov.

6.5.3.2.3 *Flamenco+ Setup Guide*

A previously developed guide to setting up Flamenco+ is available upon request. Please contact the NESC at NESC@nasa.gov.

7.0 Analysis Results

7.1 Results of Discipline Analysis

The NESC assessment data team performed initial search and analysis for several systems as requested by a subset of the discipline experts. Initial search and analysis means that the team applied the data tools to the data sets for specific ISS subsystems or problem sets and extracted what appeared to be trends and/or significant anomalies. Determinations of significance are left to the discipline experts. The trends may or may not be significant, and the anomalies may be significant but may turn out to be well understood and previously dispositioned.

The following ISS subsystems (or discipline areas) had some initial search and analysis performed: Environmental Control and Life Support Systems (ECLSS), mechanisms, software, electrical power, and human factors. Brief summaries of each are provided below. Some of the search and analysis was performed broadly using standard and enhanced search techniques, where the focus was not necessarily to capture every nonconformance related to a particular issue. In other cases, the NESC assessment team was asked to examine a specific issue and performed a deeper dive (i.e., a more exhaustive search for focused areas), such as ECLSS and mechanisms. For these cases, search, enhanced search, and statistical text mining were used.

EMU: The NESC assessment team was asked to search for anomalies related to the EMU fan bearings, meaning any nonconformances against the fan, pump, and separator bearings. Dating back to 1979, 44 related nonconformances were identified in the GFE PRACA, MOD AR, and PART IFI databases. Seven were identified as “possibly” interesting, 26 as “probably” interesting, 10 as “definitely” interesting, and 1 as not interesting. This deep dive utilized SAS® standard and enhanced search to improve the likelihood that all related anomalies were identified. The data are available upon request. Please contact the NESC at NESC@nasa.gov.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		33 of 48	

Mechanisms: The NESC assessment team was asked to search for anomalies related to peristaltic and harmonic drive pumps. Tableau® was used to perform search and the information was provided to the mechanisms discipline experts.

Software: The NESC assessment team worked with the NASA Technical Fellow for Software to identify trends related to software anomalies. The NASA Technical Fellow for Software was looking for supporting data to define the state of the discipline across the Agency. For example, trends as seen in Figure 7.1-1 were provided for use. This figure identifies failures related to ISS computers over the 5-year period from 2009 through 2014. Additional information on software failures can be obtained by contacting the NESC at NESC@nasa.gov.

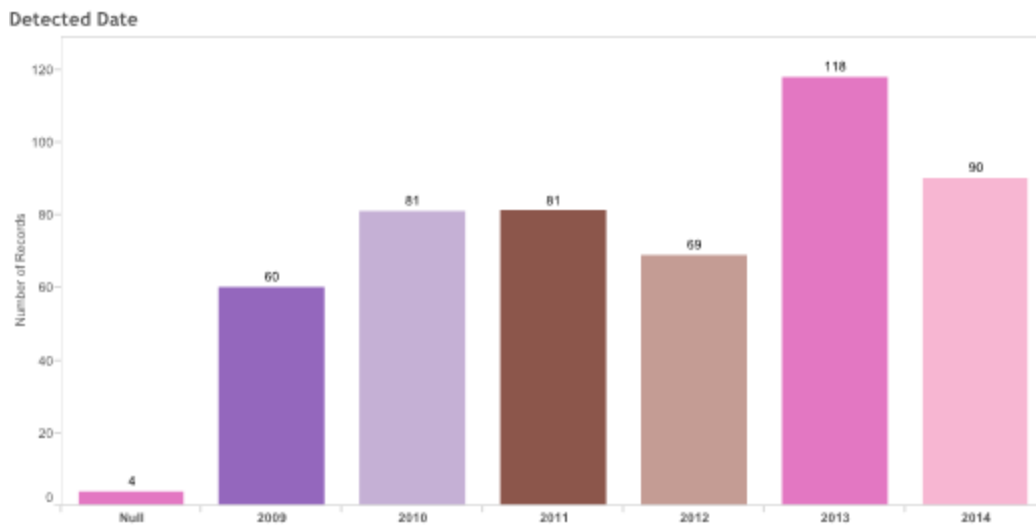


Figure 7.1-1. Trends of “ISS Computers” Failures from 2009 to 2014

Human Factors: The NESC assessment data team also worked with the NASA Technical Fellow for Human Factors and the Human Factors TDT Deputy to provide high-level trends for consideration. For instance, some of the identified trends indicated areas where astronauts are doing repeated work. These might benefit from improvements in processes instead of technical fixes. Another example can be seen in Figure 7.1-2. Nonconformances with either “smoke” or “fire” and “alarm” show an increasing trend both over an 11-year and a 5-year period using a quadratic trend curve fit, as seen in Figure 7.1-2. The human factors team may consider whether the recent uptick is significant and whether any actions are warranted.

YR	Anomalies
4	27
5	14
6	8
7	13
8	8
9	7
10	3
11	6
12	3
13	9
14	10

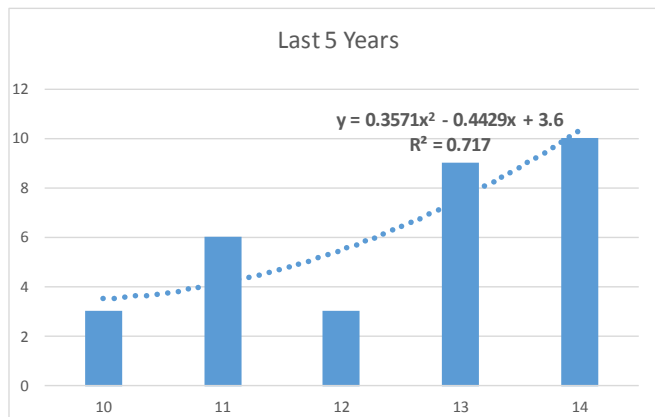
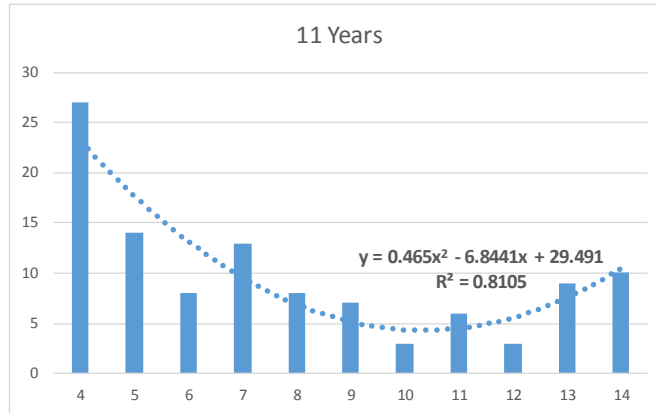



Figure 7.1-2. Nonconformances Containing “Smoke” or “Fire” and “Alarm”

Electrical Power: SAS[®] data and text mining tools were utilized to begin investigating failure trends in electrical power. The electrical power subsystem was a test case for developing a process using SAS[®] text mining that might be applied to other ISS subsystems analysis. This effort was not completed and may or may not prove beneficial. Additional explanation is provided in Appendix G. Results are not ready to be reported at this time.

Tool Suite Results: Simply using search on merged data sets added value compared with searching nonconformance databases separately. The same results could have been achieved by combining these search results using the latter approach; however, that would have been considerably more cumbersome and time consuming.

The merged data set improves the ability to trend across the broader data set, and Tableau[®] makes it straightforward and intuitive to view and parse the data. This allows the users to investigate counts and trends, as well as perform data exploration. However, some overlap between the data sets often skews the counts. For instance, at times a nonconformance identified in the MOD AR data set results in a nonconformance in the PART IFI data set, which may then end up in one of the GFE data sets. This adds burden to the user to manually remove duplicates, once identified.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		35 of 48	

Tableau[®] has proven useful for search and discovery. It is also valuable for exporting data/information to other tools, such as Excel[®].

7.2 Data Enrichment Results

Concept tags were added to the merged data set near the end of the assessment and were not utilized enough to test their efficacy. They are incorporated in Flamenco+ and Tableau[®] and, at the very least, will improve the ability to perform deep dives on particular topics where a search needs to be as comprehensive as possible. Given this, it is also expected that the concept tags will help improve the overall speed and accuracy of performing searches in general.

Proxy codes were also added to the merged data set near the end of the assessment. Testing was performed on the proxy codes, and limitations were identified that primarily stemmed from the inadequacy of the existing manual condition codes found in the existing data sets (i.e., GFE and PART PRACA). Significant manual coding errors were found during the process of developing the proxy code rules. Although the team attempted to overcome this, preliminary estimates of proxy code recall were about 30 percent.

Flamenco+, an open-source search and visualization tool for multidimensional search, was customized to use the hierarchical indexes provided by STAT and the Aerospace Ontology for the data sets. Flamenco+ was also adapted for evaluating codes. Corresponding STAT adaptations were made to provide output to support use of Flamenco+ for evaluation of proxy codes. Due to the late incorporation of proxy tags in the merged data set, Flamenco was not utilized for search during this assessment but is available for use going forward. Flamenco should be optimized for search using concept tags. Integrated use of Flamenco+ and Tableau[®] was not explored but is feasible and promising.


SAS[®] was used to perform statistical text mining on the merged data sets, focusing on specific subsystems and/or classes of anomalies. This was partially successful. This approach proved useful for identifying potential areas of interest by grouping similar anomaly topics. However, since the methodology used in SAS[®] is statistically based on word frequency, many of the clusters of anomalies identified turned out to be uninteresting. Consequently, wading through identified clusters is time consuming and often uninformative.

7.3 Topic of Interest

7.3.1 Relating System Hazards and Causes with Problem or Anomaly Occurrences

NASA S&MA organizations desire the ability to associate a potential issue or hazard to the historical operational anomalies or failures that could have led to the realization of the hazard. A system or operational hazard is defined as a risk condition that arises during operation(s) that can potentially lead to a loss of assets, mission, or personnel. Associating operational anomalies with those risk conditions can aid in understanding how those risks develop during operations and lead to better ways to prevent their development.

In the vast majority of documented cases, the occurrence of an anomaly or failure does not ultimately lead to a catastrophic consequence described by a hazard. However, it is logical to

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP-14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 36 of 48	

conclude that the occurrence of anomalies or failures during system operation should be related to the likelihood of occurrence of accidents or mishaps that result in the realization of a hazard (i.e., loss of assets or personnel). That is, documented occurrences of anomaly incidents such as those contained in the merged data set described in this study may be used to identify “close calls” or “precursors” to future catastrophic events.

This section describes a methodology to use the ISS anomaly data sets and search capabilities described in this study to identify and cluster for further analysis the anomalies associated with individual ISS hazards.

7.3.1.1 Use Case Objective

This objective is to provide a way to relate anomaly record information from the study’s ISS merged anomaly data set to potential risks and hazards defined in the ISS Hazard Analysis System.

A hazard defines a potential risk/mishap that can occur during operation(s). Within NASA, system hazards are described through the use of hazard analyses and reports, with underlying standardized hazard description wording. Given a hazard of concern or interest, it is desired to develop a compilation of the historical anomalies that have occurred that could have led to the occurrence of the hazard, and potentially rank the hazard risk by the number of related anomaly counts. (Related anomalies can be regarded as “precursors” to individual hazard occurrences.)

7.3.1.2 Method

ISS hazard information is currently available in the NASA ISS Hazard Data System, and user access to that system will be necessary to obtain the hazard information. The system allows access to ISS hazard reports in portable document format, so detailed information about individual hazards must be manually obtained by reading the reports. Figure 7.3.1-1 shows the search page image associated with the ISS Hazard Data System that can be used to retrieve specific hazard analysis reports. The user may search for hazard reports on subsystems/ payloads, hardware categories, or several other fields. For instance, if a user is interested in the “Hazard” record type associated with the “ECLSS (Environmental Control and Life Support Subsystem)” payload for the “Assembly Complete (AC)” ISS flight applicability, the user would select the options as shown in Figure 7.3.1-1. The NASA Hazard Data System will then retrieve the relevant hazard report files. Once a desired hazard report is retrieved, the report will need to be read to extract the relevant information to search for related anomalies.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		37 of 48	

Words

Record IDs
(comma-separated)

Status

 INWORK
 REVIEW
 PHASE I COMPLETE
 PHASE II COMPLETE
 PHASE III APPROVED
0 selected

Report Numbers
(comma-separated)

Subsystem/Payload

Hardware Provider
Filter Hardware Provider

Record Type

 Safety Data Package
 Hazard
 Reflight 622
 NCR
1 selected

Hardware Category

 CFE
 GFE
 Integration
 IP Caroo
0 selected

Document Type

 Applicability Matrix
 DTO
 Meeting
 MHA
0 selected

Flight of Record
Filter Flight of Record


Flight Applicability

 OS
 9A.1
 9P
 9S
 AC
1 selected

Figure 7.3.1-1. NASA ISS Hazard Data System Search Page

The primary information needed from the hazard reports is a description of the hazard causes and, perhaps, the associated controls. In many cases, a hazard cause, as stated, is analogous to a failure mode of an item or component that can lead to the realization of the hazard. In other cases, the hazard control section will identify the items or components whose failure jeopardizes the prevention of the hazardous event. This combination of a cause/failure mode with the associated component can then be used to map into the integrated anomaly database search capability.

From the hazard cause statements and/or the hazard cause control statements, a specific system component or item should be described, whose anomalous behavior can be attributed as potentially causing the hazard to occur. Relevant statements will have a syntactic form or phrase such as: a “failure (of some type or mode)” of a “component or item” during some operation can lead to the occurrence of the hazard.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		38 of 48	

The use case objective here is to use these identified component and failure characteristics to find a related set of recorded anomalies from the integrated ISS database using the search and retrieval tools developed during this study.

Example

In this example, the hazard of interest or concern is the ISS hazard report with the title “IVA Crewmember Exposure to Inadequate Respirable Atmosphere,”¹ with the associated hazard condition description of “Failure to maintain atmosphere partial pressure of oxygen and nitrogen within proper limits resulting in personnel injury/death.”


The report identifies three associated causes for the hazard:

- **Cause 1.** Low partial pressure of oxygen due to crew metabolic usage.
- **Cause 2.** Leakage/rupture of nitrogen distribution/transfer system.
- **Cause 3.** Inadvertent/excessive nitrogen introduction or release through the nitrogen pressure relief valve.

Note that causes 2 and 3 already have the structure of “failure mode” of some “component or item.” However, further descriptions of components and failure modes are found in the Controls section of the hazard report. The following cause control descriptions are excerpted from the hazard report:

- **Cause 1 Controls:** Intermodule ventilation will be established at the beginning of each ingress activity, by fans, and ducting between the Service Module (SM), Functional Cargo Block, pressurized mating adapters, Node 1, United States (U.S.) Laboratory, the airlock, and the orbiter. Control of oxygen levels will be performed by either the orbiter, while open to the station, or the SM. After orbiter departure, the airlock and the U.S. Lab will provide control of oxygen levels, introducing oxygen by use of a high-pressure oxygen tank external to the airlock and a pressure control assembly (PCA), which introduces oxygen into their volume via an oxygen introduction valve (OIV). The Inter-module Ventilation disburses oxygen throughout the ISS.
- **Cause 2 Controls:** The United States On-orbit Segment Nitrogen Distribution System is composed of three subsystems: Supply, Recharge, and Low Pressure Distribution. The nitrogen system components (i.e., recharge and distribution) are designed with either metal-to-metal or dual O-ring seals at joint interfaces (i.e., quick disconnects or gamah fittings). A single elastomer seal exists in the PCA nitrogen introduction valve (NIV).
- **Cause 3 Controls:** PCA NIVs, located in the U.S. Laboratory, and the airlock are initialized closed and normally remain in the closed position. Each PCA can be configured to automatically introduce nitrogen based on the total cabin pressure measured by the cabin pressure sensor. In the automatic mode, the NIV valves will be commanded open if the total cabin pressure falls below a threshold. The NIVs will

¹ The hazard number for this example is ISS-ECL-0206-AC.

	NASA Engineering and Safety Center Technical Assessment Report	Document #: NESC-RP- 14-00950	Version: 1.0
Title: ISS Anomalies Trending Study		Page #: 39 of 48	

remain open until the threshold is reached. Also, each NIV can be manually opened or closed by the crew, or remotely commanded by the crew/ground. NIVs remain in the last commanded position. The PCA is a “must work” function.

Cursory review of the wording in the Cause or Controls sections can identify several potential components/items that are important to the system operations and that contribute to inhibiting the hazard occurrence. Selected entities are:

- PCA
- OIV
- NIV
- Nitrogen distribution/transfer system
- Nitrogen pressure relief valve
- Cabin pressure sensor
- High-pressure oxygen tanks

In addition, various failure modes identified include:

- Low [partial] pressure
- Leakage/rupture
- Inadvertent/excessive [gas] introduction or release

The analyst or engineer involved in this process should also have the system knowledge to elicit or infer other component/items and failure modes/causes for review purposes.

Using the Tableau® Search Capability

Searches are performed using the Tableau® capability to access the integrated ISS data set, based on the context described above. For example, Figure 7.3.1-2 shows the main Tableau® search screen that results with the terms *pca*, *oiv*, and *niv* used as search parameters.

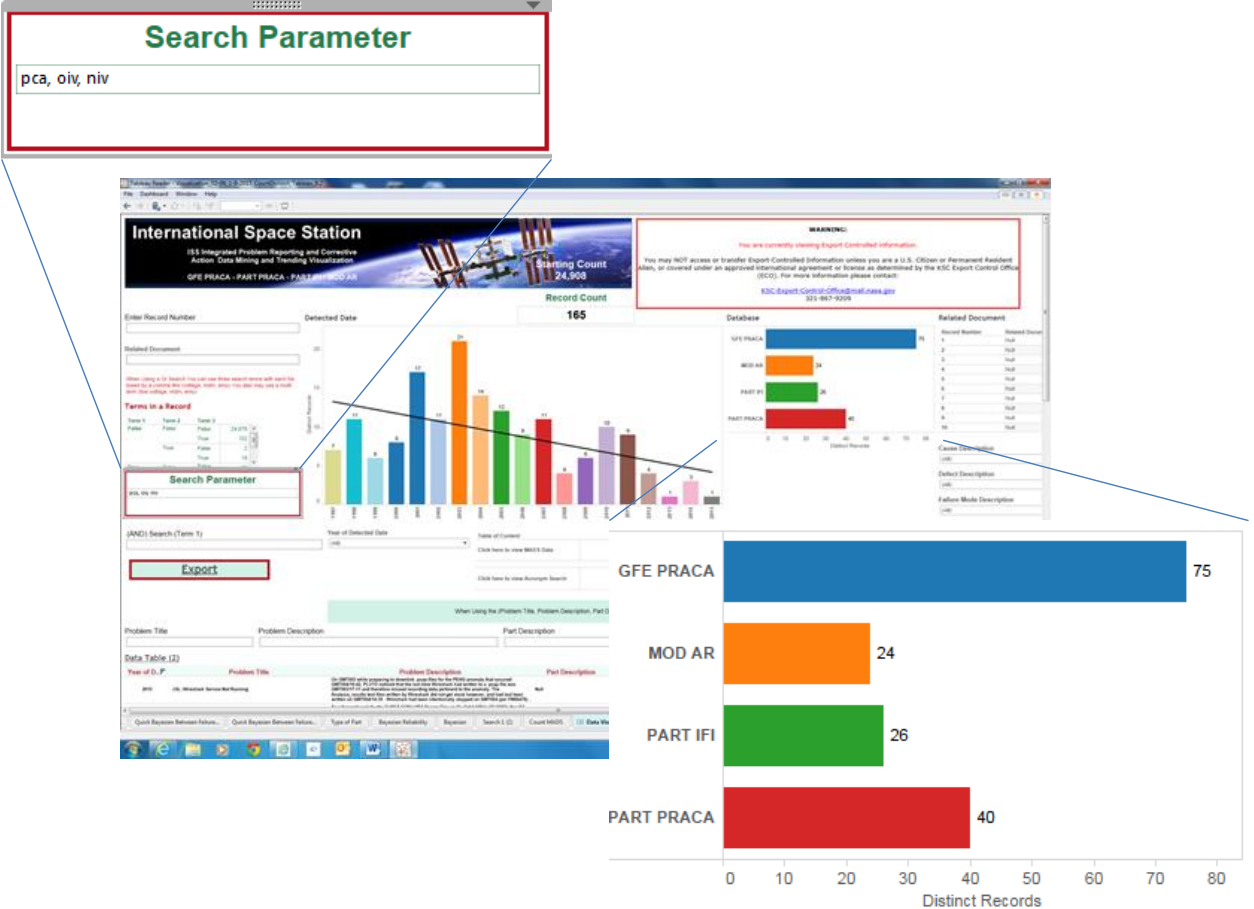


Figure 7.3.1-2. Tableau® Search Screen with Three Search Parameters

In this case, the search retrieved 165 records that contained one of the three search parameters, 40 of which were from the PART PRACA data set, 26 from the IFI data set, 24 from the MOD AR data set, and 75 from the GFE PRACA data set. A selected portion of the anomalies with titles and descriptions that are related to the three components pointed to by the hazard report is shown in Figure 7.3.1-3.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		ISS Anomalies Trending Study	
		Page #: 41 of 48	


Low Pressure O2 Regulator Internal Leakage	<p>On about GMT 200 04:00 after Oxygen tank installation once the Low and High Pressure Supply Valves were opened, it was noted that the oxygen pressure from the Lab and Airlock PCA oxygen pressure sensors was steadily increasing. This was analyzed to be an internal leak through the oxygen regulator assembly (in excess of 0.025 scc/s or about 25 times above specification) of about 113kPa/day (16.4 psi/day). A/L OIV was opened causing the regulator to ..</p> <p>On about GMT 2001/200 04:00 after Oxygen tank installation once the Low and High Pressure Supply Valves were opened, it was noted that the oxygen pressure from the Lab and Airlock PCA oxygen pressure sensors was steadily increasing. This was analyzed to be an internal leak through the oxygen regulator assembly (in excess of 0.025 scc/s or about 25 times above specification) of about 113kPa/day (16.4 psi/day). A/L OIV was opened causing the regulator to ..</p>
Oxygen System Supply Line Depressurization	<p>During troubleshooting of the Low Pressure Regulator Leak on GMT 228 an unexplainable pressure signature caused a halt in troubleshooting until the pressure signature is understood. The troubleshooting proceeded as follows:1) Supply valve closed (GMT 228:00:13)2) OIVs in Lab and Airlock opened to reduce line pressure to ambient (GMT 228:00:15:00)3) Supply valve opened to "slam" regulator (GMT 228:00:20)4) System allowed to flow for 10 minutes 5) OIVs cl..</p>
PCA Cabin Pressure Sensors Out of Spec	<p>The Lab and Airlock PCAs total pressure reading exceed 0.02 psi (0.01 accuracy for each PCA, for a worst case difference of 0.02 psi) when the Station is equalized between the two. At Airlock activation the difference was about 0.04 psi. The probably cause of this is helium contamination. Back in early 2001, the Airlock was leak tested. This testing included the addition of a certain amount of helium into the cabin atmosphere of the Airlock. Nominally the PCP (Pressure Contr..</p>
PCA Change from Monitor to Safe Mode	<p>For moving the Soyuz vehicle from the SM aft docking position to the FGB TK port, the Station was configured as follows: Both Lab hatches were closed, and the Lab aft port and stbd IMV valves were open. All (6) Node 1 hatches were closed, and the aft port and forward port and stbd IMV valves were open. The FGB hatches were closed. The SM hatches, except the PKOIRO hatch, were closed. The Soyuz undocked from the SM aft position and at GMT 2001/055:10:37 docked a..</p>
Unexpected PCA Inlet Pressure Alarms	<p>During the initial purges N2 system, a PCA Low N2 Pressure high alarm at 127 psia was set. The alarm is set to 120 psia and the maximum spec lockup pressure is 145 psia (with a 14.7 cabin) of the pressure regulators. When we are not flowing then the pressure could rise above alarm with in spec regulator performance. This also applies to the O2 system and at all PCAs.</p>

Figure 7.3.1-3. Anomaly Text Information Results for Associated Hazard Components/Items

The associated failure mode descriptions are also provided by the Tableau® search and are shown in Figure 7.3.1-4. The failure modes of interest from the hazard report deal with low pressure, leakage/rupture, or inadvertent release. From the information in Figure 7.3.1-4, several of the 24 records with descriptions such as EXTERNAL/INTERNAL LEAKAGE, STRUCTURE FAILURE, or PREMATURE OUTPUT map to these kinds of failure modes, and the user may select these particular records for more detailed examination to determine how closely the anomalies relate to the hazard conditions presented in the hazard report.

Failure Mode Description	Count
Fails Open or Fails to Close (Retract) Completely	2
Out of Tolerance (Function)	1
Not Applicable	1
Temp / Pressure High	1
Other	1
Improper Installation	2
STRUCTURE FAILURE	1
EXTERNAL LEAKAGE	2
Incorrect Part	1
UNSATISFACTORY CONDITION	1
Hardware Not Per Drawing	2
INTERNAL LEAKAGE	1
Hardware Not Per Drawing	1
Intermittent Operation	2
FAILS ON	1
Premature Inadvertent Output (Operation) or Shutdown	1

Figure 7.3.1-4. Failure Mode Descriptions Associated with Identified Anomaly Records

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		42 of 48	

As an additional source of information, the associated part information resulting from the Tableau® search is shown in Figure 7.3.1-5. This part information helps to corroborate that the components/items associated with the anomalies are those of interest (i.e., those found in the hazard report).

Part Number	Part Description	Database Name	
683-16421-6	LOW PR O2 REG/RLF VLV ASSY	PART PRACA	1
2353052-1-1	Oxygen/Nitrogen Isolation Valve	PART PRACA	4
	OXYGEN/NITROGEN ISOLATIONVALVE	PART PRACA	1

Figure 7.3.1-5. Part Numbers and Descriptions Associated with Identified Anomaly Records

Observations

The methodology discussed above provides a tool for S&MA personnel, as well as other interested organizations, to identify incidents that have occurred in the past that could have led to a critical or catastrophic mishap or event. These incidents may be reviewed, counted, and trended to raise awareness and assess whether preventive actions would be prudent. The ability to search across several databases to identify the appropriate incidents is a key attribute to finding a more complete set of incidents for analysis.

7.4 Description of Future Analysis Plans


The original plan called for the NESC assessment team to identify ISS trends and/or significant anomalies. This work was not completed, due largely to the cleanup, merging, and data-mining efforts being more challenging and time consuming than expected. The assessment lead will work through the Systems Engineering TDT and the NESC Review Board to develop a plan going forward.

8.0 Findings, Observations, and NESC Recommendations

8.1 Findings

The following findings were identified:


- F-1.** The expected goals and outcomes of the data-mining effort determine which data sets and fields are required. For example, importing problem descriptions was essential for performing problem trends. However, disposition and corrective action text fields, where available, would likely have been helpful but were not carried over to the merged data set and, therefore, were not available for trend analysis.
- F-2.** On occasion, searching the merged data set can result in over-counting frequencies and trends. PART IFIs are often elevated and repeated in PART PRACAs. On occasion, AR records with reoccurrences of a problem in a single data record can result in under-counting frequencies and trends.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		43 of 48	

- F-3.** Proxy code development efforts that were based on manual trend codes proved not to be effective because too many errors were found in the manual trend codes.
- F-4.** Visualization tools can be successfully customized for querying and filtering merged data set and displaying query results in multiple displays for the user.
 - Demonstrated with Tableau® and Flamenco.
- F-5.** The tool suite developed in this assessment showed promise in supporting discipline experts in performing deep investigations into technical issues.

8.2 Observations


- O-1.** The data analysis team demonstrated the ability to create a searchable, merged problem data set from multiple problem reporting systems by overcoming problems/limitations between fields and dissimilar field values for individual data sets.
- O-2.** Concept tags based on modifications to the Aerospace Ontology were created for all of the records in the merged data set. These tags were integrated into the merged data set late in the assessment and were not fully evaluated.
- O-3.** Within the Tableau® Desktop framework, the merged data set may have reached its performance limits, so that expanding the number of records, faceted search, or visualization capabilities will require server-based systems.
- O-4.** Standard query-type searches are limited in that they will not catch multiple synonyms, alternate spellings, abbreviations, and acronyms.
- O-5.** Complexity in user interfaces for search requires users to have additional training to maximize the benefits of information retrieval.
- O-6.** The assessment was not able to fully realize strategies for information retrieval based on multidimensional faceted search.
- O-7.** STAT strategies for tagging of complex phrases sometimes obscure properties that are important search terms. Words such as “inadvertent” are merged into a generic “bad” set of variants applied to operations. The resulting concept tag emphasizes the type of operation/function rather than the type of problem.
- O-8.** Lexical analysis and text filtering can be refined so that a second round of data cleaning is avoided during review of candidate words and phrases for the Aerospace Ontology vocabulary.
- O-9.** The SAS text-mining process can be redesigned for improved recall and precision by including concept tags.
- O-10.** It was demonstrated that anomaly record information from the ISS merged data set can be related to potential risks and hazards defined in the ISS Hazard Analysis System.

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		44 of 48	

8.3 NESC Recommendations

The following NESC recommendations are directed to future users or implementers of this tool suite, or to developers who will merge and trend across multiple data sets. These recommendations are intended to achieve a robust tool suite framework and data set for analyses of anomaly groups and trends.

- R-1.** In future data mining efforts across multiple reporting systems, carefully align the objectives and expected outcomes of the investigation with the selected problem reporting systems and their reporting processes, recognizing the possibility of duplicate records. *(F-1)*
- R-2.** To perform more accurate problem counts and trends across the PART and AR data sets, develop methods and capabilities to aid the user in merging, associating, or eliminating duplication to support the goals of the trending. *(F-2)*
- R-3.** The Agency should develop a minimal set of common data fields and field values that are clearly defined for use in problem reporting data sets. *(O-1)*
- R-4.** Consider using query reformulation. Variant lists can be included in the user interface so that if one of the words or phrases in the list is entered as a search term, others in the list can be offered. The user can review these and build a better query. Updates to the Aerospace Ontology should include additional variants for these data sets. *(O-4)*
- R-5.** Integrate the concept tags from STAT/Aerospace Ontology into information retrieval strategies in the search and visualization tools and evaluate their effectiveness. *(F-3, O-9)*
- R-6.** Explore strategies where faceted search uses hierarchy in the data to look ahead and filter and, thus, complements search and filtering in the visualization tool. *(O-6)*
- R-7.** Improve processing of complex expressions in text and use a negative properties facet to provide better indexing of types of problems. *(O-7)*
- R-8.** Develop look-ahead strategies with dimensional partitions (facets) for quick browsing, summaries, conceptual metadata, and accessible information on the types of data in each data source. These dimensions should be specified to highlight common features of nonconformances. *(O-6)*
- R-9.** Investigate further, with the ISS S&MA community, the use of the merged data set and tool suite developed during this assessment to gain a better understanding how past ISS operations reflect on existing ISS hazards. *(O-10)*

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		45 of 48	

9.0 Alternate Viewpoint

There were no alternate viewpoints identified during the course of this assessment by the NESC team or the NRB quorum.

10.0 Other Deliverables

In addition to this final report of findings, observations, and associated recommendations regarding ISS significant anomalies and/or trends, the following deliverables were provided to the stakeholders:

- The current ISS anomaly data set, accessible by way of a tool suite, to include the graphical user interface.
- Training and reference material documenting lessons learned and configuration of the tool suite to support any future trending activities beyond ISS.

11.0 Lessons Learned

11.1 Preventing Errors in Problem Reporting Codes


11.1.1 Description

Fields for manually assigning problem reporting codes were included in some of the databases in the ISS anomaly data set. The coding schemes for types of failure modes and defects produced coding errors and made search by codes less effective. The coding errors were discovered while designing rules for generating proxies for these codes based on content in the text fields in the reports. In the GFE PRACA and PART PRACA data sets, manual coding errors were much worse than expected.

Multiple possible types of coding errors can occur:

- Misinterprets code definitions (Help text) or is unable to fill in gaps in short definitions.
- Misinterprets how to assign codes to multiple condition fields, especially when there is some overlap.
- Misinterprets text description in report or cannot guess missing information in the report.
- Chooses a nonspecific code.
 - Varying reluctance to commit to specific code.
 - Appropriate code not found in set.
- Uses only a subset of codes to handle difficult coding schemes.
- Copies a code from a related report (which may be incorrect).

Many problem reporting codes are not clearly defined. The definitions (Help text) are brief and confusing. No guidance is given on what code assignment should be used when multiple alternative codes are possible. Data overload for users occurs because the code sets are large and

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		46 of 48	

multilayered, with complex, inconsistently structured fields and codes. Types of relations between the fields are not explicit or well-defined. Subtype-super-type relations are mixed with other relations in code hierarchies, violating the assumption that all the characteristics of the superset are applicable for the members of the subset.

11.1.2 Corrective and/or Preventive Actions

Procedures for developing and reviewing coding schemes should be defined, with emphasis on clarity and ease of use by both coders and analysts. Codes and problem reporting fields need to be well-defined and distinct. Criteria for assigning each code need to be expressed in definitions that are long enough for clarity, with sufficient examples and detail. They should be expressed in terms that are aligned with the language used in the text fields of the reports. If the coder is constrained to select a single code and no secondary codes are allowed, then guidance is needed as to what characteristics should be primary or preferred in assigning the code. This information should also be available to analysts who use the codes to retrieve records.


Coding schemes should be evaluated by inter-rater reliability studies before they are released. Reproducibility is frequently measured as inter-rater reliability between two or more coders. Code selections should be regularly reviewed, and coding errors should be corrected. Results of the reviews should be used for updating coding schemes and definitions. Systems for training and help should be provided, such as advice and additional information in FAQs.

12.0 Recommendations for NASA Standards and Specifications

No recommendations for NASA standards and specifications were identified as a result of this assessment.

13.0 Definition of Terms

Corrective Actions	Changes to design processes, work instructions, workmanship practices, training, inspections, tests, procedures, specifications, drawings, tools, equipment, facilities, resources, or material that result in preventing, minimizing, or limiting the potential for recurrence of a problem.
Finding	A relevant factual conclusion and/or issue that is within the assessment scope and that the team has rigorously based on data from their independent analyses, tests, inspections, and/or reviews of technical documentation.
Lessons Learned	Knowledge, understanding, or conclusive insight gained by experience that may benefit other current or future NASA programs and projects. The experience may be positive, as in a successful test or mission, or negative, as in a mishap or failure.
Observation	A noteworthy fact, issue, and/or risk, which may not be directly within the assessment scope, but could generate a separate issue or concern if not addressed. Alternatively, an observation can be a positive


	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		Page #:	
ISS Anomalies Trending Study		47 of 48	

acknowledgement of a Center/Program/Project/Organization's operational structure, tools, and/or support provided.

Problem	The subject of the independent technical assessment.
Proximate Cause	The event(s) that occurred, including any condition(s) that existed immediately before the undesired outcome, directly resulted in its occurrence and, if eliminated or modified, would have prevented the undesired outcome.
Recommendation	A proposed measurable stakeholder action directly supported by specific Finding(s) and/or Observation(s) that will correct or mitigate an identified issue or risk.
Root Cause	One of multiple factors (events, conditions, or organizational factors) that contributed to or created the proximate cause and subsequent undesired outcome and, if eliminated or modified, would have prevented the undesired outcome. Typically, multiple root causes contribute to an undesired outcome.
Supporting Narrative	A paragraph, or section, in an NESC final report that provides the detailed explanation of a succinctly worded finding or observation. For example, the logical deduction that led to a finding or observation; descriptions of assumptions, exceptions, clarifications, and boundary conditions. Avoid squeezing all of this information into a finding or observation.

14.0 Acronym List

AMA	Analytical Mechanics Association, Inc.
AR	Anomaly Report
DR	Discrepancy Report
EG	Enterprise Guide
EMU	Extravehicular Mobility Unit
ETL	Extract, Transform, and Load
EVA	Extravehicular Activity
GFE	Government-furnished Equipment
IFI	Items for Investigation
ISS	International Space Station
JSC	Johnson Space Center
KSC	Kennedy Space Center
LaRC	Langley Research Center
MADS	Maintenance Analysis Data Set
MOD	Mission Operations Directorate
MTSO	Management and Technical Support Office
NESC	NASA Engineering and Safety Center

	NASA Engineering and Safety Center Technical Assessment Report	Document #:	Version:
		NESC-RP-14-00950	1.0
Title:		ISS Anomalies Trending Study	
		Page #: 48 of 48	

NGO	Needs, Goals, and Objectives
NIV	Nitrogen Introduction Valve
NRB	NESC Review Board
OIV	Oxygen Introduction Valve
PART	Problem Analysis Resolution Tool
PCA	Pressure Control Assembly
PRACA	Problem Reporting and Corrective Action
QARC	Quality Assurance Record Center
S&MA	Safety and Mission Assurance
SCR	Software Change Request
SEO	Systems Engineering Office
SM	Service Module
STAT	Semantic Text Analysis Tool
TDT	Technical Discipline Team
U.S.	United States

15.0 References

1. “Space Shuttle and International Space Station Recurring Anomalies,” NASA Engineering and Safety Center RP-05-10, January 19, 2005.
2. Malin, J. T., Millward, C., Schwarz, H. A., Gomez, F., and Throop, D.: “Semantic Annotation of Aerospace Problem Reports to Support Text Mining,” *IEEE Intelligent Systems*, Vol. 25, No. 5, September/October 2010, pp. 20–26.
3. Morville, P. and Callender, J.: *Search Patterns: Design for Discovery*, O’Reilly, 2010.

16.0 Appendices (separate volume)

- Appendix A. Outline of Concept of Operations (ConOps)—International Space Station (ISS) Anomalies Trending Study
- Appendix B. Lexical Analysis of the Text in Anomaly Reports
- Appendix C. Semi-Automated Ontology Updating from Corpus Analysis Results
- Appendix D. Basic Process for Customizing and Updating the Aerospace Ontology
- Appendix E. Data Visualization
- Appendix F. Refining Proxy Codes
- Appendix G. SAS[®] Analysis with Text Mining Topics
- Appendix H. ISS Data Mining Site Construction Guide

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 01-12-2015		2. REPORT TYPE Technical Memorandum		3. DATES COVERED (From - To) May 2014 - September 2015	
4. TITLE AND SUBTITLE International Space Station (ISS) Anomalies Trending Study				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Beil, Robert J.; Brady, Timothy K.; Foster, Delmar C.; Graber, Robert R.; Malin, Jane T.; Thornesbery, Carroll G.; Throop, David R.				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 869021.01.07.01.01	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Langley Research Center Hampton, VA 23681-2199				8. PERFORMING ORGANIZATION REPORT NUMBER L-20648 NESC-RP-14-00950	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001				10. SPONSOR/MONITOR'S ACRONYM(S) NASA	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) NASA/TM-2015-218991/Volume I	
12. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 59 Mathematical and Computer Sciences (GENERAL) Availability: NASA STI Program (757) 864-9658					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The NASA Engineering and Safety Center (NESC) set out to utilize data mining and trending techniques to review the anomaly history of the International Space Station (ISS) and provide tools for discipline experts not involved with the ISS Program to search anomaly data to aid in identification of areas that may warrant further investigation. Additionally, the assessment team aimed to develop an approach and skillset for integrating data sets, with the intent of providing an enriched data set for discipline experts to investigate that is easier to navigate, particularly in light of ISS aging and the plan to extend its life into the late 2020s. This report contains the outcome of the NESC Assessment.					
15. SUBJECT TERMS Data Mining; Anomaly; International Space Station; NASA Engineering and Safety Center					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
U	U	U	UU	53	19b. TELEPHONE NUMBER (Include area code) (443) 757-5802