Observatory Performance Metrics Best Practices

A white paper from the observatory best practices/lessons learned series

Thomas Kearney
Leslie M. Smith
Christopher Rutherford

June 30, 2019

Table of Contents

Executive Summary

The focus of this paper is effective performance metrics development and how these metrics can improve management success. As best practices implementation objectives are defined, performance metrics can be developed to manage and measure progress toward best practice implementation. Performance metrics are developed within a hierarchy of observatory strategic long term and short term goals. This paper also provides example metrics for four observatory topic areas as a means to stimulate discussion of how metrics can be utilized in real-life scenarios.

Observatory performance metrics are developed, applied, managed and maintained within the context of overall observatory's mission, goals, objectives and capabilities. Generally, for an observatory these are described in the Science Plan, Concept of Operations Plan (ConOps), and Annual Work Plan (AWP). A typical tool to use for a performance metric is the Key Performance Indicator (KPI) that describes a measurable indicator that demonstrates how effectively a company is achieving key business objectives. In other words, the goal of performance KPIs and their associated metrics is to provide measurable indicators of success. There will always be a need to revisit performance measure design to ensure alignment with the observatory's mission, goals and objectives.

In order to assess the state of the industry in terms of performance metrics, literature review research was conducted and effective performance metrics developed and managed within a hierarchy of business management practices were reviewed. Each of these best practices are discussed in detail, accompanied by context and literature references in the remainder of the white paper. Additionally, these best practices have been organized into a best practice Self-Assessment Tool that enables an existing or new organization to assess their current data identification, citation and tracking capabilities and maturity level (See Appendix).

Best practices described in this white paper are based on an extensive survey of existing observatory best practices. They represent an idealized world of achievable best practices, which are recognized to be challenging to implement. Each observatory has its own priorities and available resources, as such, the best practices described are aspirational. This best practice white paper objective is to provide a simplified, easy to understand and apply guide for self-assessment and planning. It does not represent a guide for technical assessments or implementation.

BP 1: Develop and maintain a Science Plan and ConOps strategic plan.
The Science Plan and ConOps strategic plan documents define the observatory mission, goals, objectives, and enabling capabilities. These provide important, foundational documents for the development of performance metrics as they provide sources of information for defining these indicators and metrics.

BP 2: Develop and maintain 1 year Annual Work Plan (AWP).
The AWP describes what the observatory expects to accomplish in the coming fiscal year. The AWP should include a series of high level performance goals (clear and agreed upon goals and objectives, performance metrics and, where appropriate, performance targets) for the coming year. The goals should include both scientific and operational issues.

BP 3: Develop and maintain Key Performance Indicators (KPIs).
A Key Performance Indicator (KPI) describes a measurable indicator that demonstrates how effectively a company is achieving key business objectives. Organizations use KPIs to evaluate and report their progress and success of reaching key objectives. When designing KPIs it is important to clarify the audience, describe its relevance to observatory goals and objectives, determine the number, grouping, and level of KPIs to report, and reference similar industry standards and benchmarking to peer groups.

BP 4: Develop and maintain Performance Metrics that are Specific, Measurable, Achievable, Relevant, Time-Bound (SMART).
A performance metric is a measurement value within a KPI that tracks performance and progress. Useful metrics have two qualities, (1) they are persistent, showing that the outcome of an action at one time will be similar to a later time; and (2) they are predictive, demonstrating a causal relationship between the action and the outcome.

BP 5: Develop, maintain, and distribute a Performance Metrics Scorecard.
Performance scorecards provide a dashboard of KPI topics and their associated performance metrics. The scorecard's objective is to quickly convey performance targets and actual results to enable effective assessments of current performance, and identify areas requiring management attention. Multiple scorecards may be needed depending on the complexity of the KPIs.

BP 6: Perform data-driven management reviews using KPIs and performance metrics.
Data-driven performance reviews are regularly scheduled, structured meetings used by organizational leaders and managers to review and analyze data on progress toward key performance goals and other management-improvement priorities. They are used to target areas where leaders want to achieve performance improvements. There is a growing body of evidence that well-run data driven management reviews lead to improved organizational performance.

BP 7: Perform performance gap analysis and continuous process improvement.
Continuous process improvement is required to ensure performance objectives and their associated measurements remain true. Gap analysis provides management with data driven information to respond with targeted corrective actions. Frequent data driven management reviews using performance scorecards provides a forum to discuss the gaps between performance targets and current actual results. Performance targets, current results and gap analysis provide information for areas of potential improvement.

Scope

This white paper on Marine Observatory Performance Metrics examines the current trends and drivers for performance metrics, identifies current industry best practices, provides best practices and a self-assessment tool.

The observatory performance metrics best practices discussed in this paper focus primarily on the methodology and structure of developing and maintaining effective performance metrics. In addition to these performance metrics best practices, specific examples of observatory performance metrics are provided. These examples are not intended to be a comprehensive list of best practices metrics, rather to stimulate ideas for observatory managers as their own metrics are defined.

These best practices have been organized into a best practice Self-Assessment Tool that enables an existing or new organization to assess their current observatory performance metrics capabilities and maturity level. This tool can also be used to identify steps to achieve the next aspirational level (See Appendix).

Background

Business Performance Metrics

Effective performance metrics are developed and managed within a hierarchy of business management practices. A brief summary of how performance metrics fit into these practices is provided for context.

In general, business management plans and performance measurement systems include:

- 3-5 year strategic plan describing long term goals
- 1 year tactical plan describing short term objectives, which include:
    - Key Performance Indicators (KPIs)
    - Performance Metrics
    - Performance Scorecard Reports

A Key Performance Indicator (KPI)[1] describes a measurable indicator that demonstrates how effectively a company is achieving key business objectives. Organizations use KPIs to evaluate and report their progress and success of reaching these key objectives.

A performance metric[2] is a measurement value within a KPI that tracks performance and progress.

[1] https://www.klipfolio.com/blog/kpi-metric-measure
[2] https://www.klipfolio.com/blog/kpi-metric-measure

A performance scorecard (Robert and Norton 1992) primarily refers to a performance management report used by a management team to evaluate and manage the implementation of a strategy or operational activities. It is a set of performance metrics that provides a broad but comprehensive view of the business objectives and associated status, with the assumption that no single metric can provide a clear indication of overall performance target success.

Examples of typical types of Business Performance KPIs and Metrics[3] include:

- Efficiency Indicators
- Effectiveness Indicators
- Capacity Indicators
- Productivity Indicators
- Quality Indicators
- Profitability Indicators
- Competitiveness Indicators
- Value Indicators

These performance management practices can also be applied to observatory management.

Performance Metrics Required for Federally Funded Observatories

Observatory performance metrics are required for federally funded programs. Federal agencies are required to develop a strategic plan, establish annual performance goals, and report on the progress made toward achieving those goals on a regular basis.

The Government Performance and Results Act of 1993 (GPRA) requires agencies to engage in performance management tasks such as setting goals, measuring results, and reporting their progress (103rd Congressional Record 1993). In order to comply with GPRA, agencies produce strategic plans, performance plans, and conduct gap analyses of projects.

The GPRA Modernization Act of 2010 took the existing requirements of the 1993 Act and developed a more efficient and modern system for government agencies to report their progress. The new law revises agency annual performance planning requirements under GPRA by requiring a link between the performance goals in the annual plan with the goals in their strategic plans (Kamensky 2011). Furthermore, the GPRA Modernization Act requires the plans to describe the strategies and resources agencies will use, and requires the plans to cover a 2-year, rather than a 1-year period.

[3] https://www.heflo.com/blog/business-management/process-performance-metrics/

In line with these federal mandates, the National Science Foundation (NSF) requires observatory performance metrics of its funded programs and facilities. Specifically, the NSF requires a set of performance goals and metrics be provided to sufficiently establish that the facility is operating successfully. Performance against these metrics are reported periodically as required by the program. (NSF 2018).

Methodology

This white paper is one of four in a series of best practice white papers. The other best practices white papers are: Data Identification, Citation and Tracking, Data Product Quality, and Community Engagement. Similar methodology was used in each of these best practices white papers.

Best Practices Research and Synthesis

Observatory Performance Metrics best practices identification, research and synthesis was an iterative building process. As best practices were identified, they were researched, refined and validated using extensive literature reviews. Once this was completed, the best practices and best practice self-assessment tools were validated through interviews with staff from three relatively mature observatories. Due to the sensitive nature of research findings, the organizations examined during research are not identified. Literature review references are included.

Our authors focused on the following research objectives while conducting secondary research:

- Determine drivers for observatory performance metrics
- Determine high-level requirements for observatory performance metrics
- Determine the current state of industry capabilities to meet drivers

Information was synthesized from this research to identify and define best practices. As needed, secondary research was revisited to refine, test, and validate best practices. The goal of this research was to provide a high level overview of the current state of the industry in implementing these best practices. This research is not meant to be a detailed technical assessment.

As best practices were identified and defined, a self-assessment tool was developed. The best practice self-assessment tool was inspired by the Capability Maturity Model (CMM) developed by the Software Engineering Institute (SEI) at Carnegie Mellon University in 1986 (Paulk et al., 1993). The self-assessment tool creates a ranking of best practices (Figure 2), providing questions and a scoring methodology. The scoring methodology provides flexibility for best practice variations across organizations.

The self-assessment tool is intended to provide a structure for internal assessment and to identify aspirational improvements that can be implemented to increase an observatory's maturity level (See Appendix). It also provides context based on current industry-wide best practice maturity levels. The tool ranking levels were validated through secondary and primary research.

Figure 1 displays one potential combination of capabilities, which result in a maturity level for a hypothetical observatory. Each observatory will have different combinations of capabilities, which aggregate to a certain maturity level. For example, one observatory may excel at _____, whereas another may excel at providing _____. A simplified capability scoring method to determine levels is described in the Appendix.

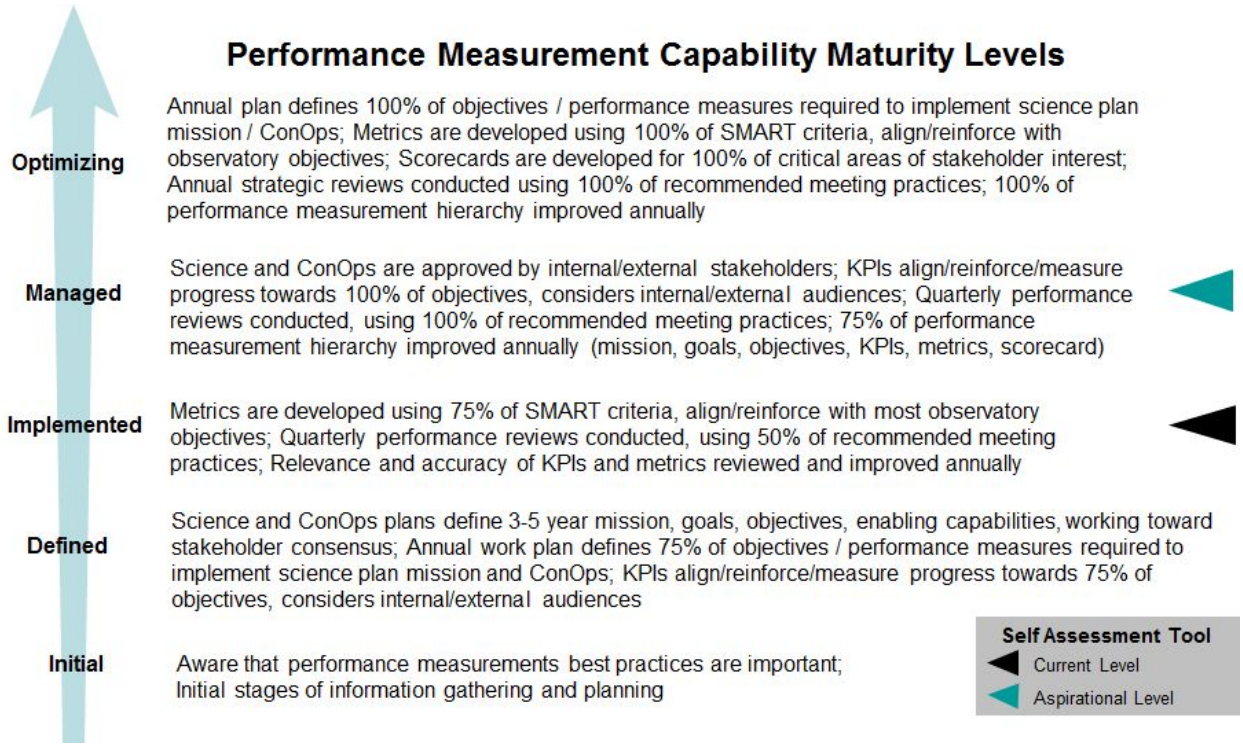

Figure 1. Best Practice Self Assessment Tool Example

Results & Discussion

Observatory Performance Metrics

Observatory performance metrics are developed and maintained within the context of an observatory's overall mission, goals, objectives and capabilities. Generally, for an observatory these are described in the Science Plan, Concept of Operations Plan (ConOps), and Annual Work Plan (AWP).

Key performance indicators (KPIs) and supporting performance metrics are developed and maintained within the context provided by the Science Plan, ConOps and AWP. Performance scorecards are KPI and metrics reports that demonstrate progress toward observatory goals and objectives.

Data-driven management reviews use KPIs, performance metrics and performance scorecards to perform performance gap analysis and inform continuous process improvement.

Observatory performance metrics best practices address each of the topics described above:

BP 1: Develop and maintain a Science Plan and ConOps strategic plan
BP 2: Develop and maintain 1 year annual work plan
BP 3: Develop and maintain Key Performance Indicators (KPIs)
BP 4: Develop and maintain Performance Metrics that are Specific, Measurable, Achievable, Relevant, Time-Bound
BP 5: Develop, maintain, and distribute a Performance Metrics Scorecard
BP 6: Perform data-driven management reviews using KPIs and performance metrics
BP 7: Perform performance gap analysis and continuous process improvement

Using these best practices to guide performance metrics development and utilizing the self-assessment tool provided, will help organizations focus on specific areas to measure their success. Implementing comprehensive business management practices, including the performance metrics described above, leverages a deep body of existing management knowledge and best practices to help observatory managers recognize opportunities for improvement and increased value to the users of their observatory information.

Observatory Performance Measurement Plans, which include KPIs and metrics, can be used to define and manage performance expectations and results. This plan should clearly align observatory objectives with performance KPIs and metrics that measure and track progress towards targeted results. Metrics may measure people, process, or technology aspects of an organization.

Science Plan and Concept of Operations

Observatory mission, goals, objectives, and enabling capabilities are typically defined in the observatory’s Science Plan and Concept of Operations Plan (CONOPS). These documents provide sources of information for defining KPIs and performance metrics.

Science Plans include the observatory’s science mission, science goals, high level science requirements, and key aspects of design to deliver on science objectives. Other qualitative aspects of a science plan may include the ability to enable transformative science, the observatory's trustworthiness, inclusiveness and reliability in the eyes of its user community.

Examples of science plans include: NEON's 2011 Science Strategy Plan - Enabling Continental-Scale Ecological Forecasting[4] and Ocean Network Canada's Strategic Plan - Understand the Ocean. Understand the Planet[5].

In general, a CONOPS is developed for operational observatories and includes operational goals and objectives required to deliver on the observatory science plan. It offers clear methodology to realize the goals and objectives for the system, while not intending to be an implementation or transition plan. CONOPS documents can be developed in many different ways, but usually share the same properties.

CONOPS typically include the following:

- Statement of the goals and objectives of the system
- Strategies, tactics, policies, and constraints affecting the system
- Organizations, activities, and interactions among participants and stakeholders
- Clear statement of responsibilities and authorities delegated
- Specific operational processes for fielding the system
- Processes for initiating, developing, maintaining, and retiring the system

BP 1: Develop and maintain a Science Plan and a ConOps strategic plan

Annual Work Plan

The Annual Work Plan (AWP) describes what the observatory expects to accomplish in the coming fiscal year. The AWP should include a series of high level performance goals (clear and agreed upon goals and objectives, performance metrics and, where appropriate, performance targets) for the coming year. The goals should include both scientific and operations issues. The goals and metrics will naturally vary from observatory to observatory. The NSF Program Officer will review the AWP goals to ensure they are aligned with the long-term scientific objectives of the facility (NSF 2018).

Goals are general guidelines that explain what you want to achieve in your community. Objectives define strategies or implementation steps to attain the identified goals. Unlike goals, objectives are specific, measurable, and have a defined completion date[6].

The AWP concept applies to all observatories, regardless of funding agency. The one year plan is guided by the observatory's Science Plan and Concept of Operations Plan in that converts the long term goals from those documents into actionable one year objectives. In addition to defining these objectives, the plan measures implementation success using key performance indicators and performance metrics.

[4] https://www.neonscience.org/sites/default/files/basic-page-files/NEON_Strategy_2011u2_0.pdf
[5] https://www.ocenanetworks.ca/sites/default/files/pdf/ONC_Strategic_Plan_2013-2018.pdf
[6] https://www.michigan.gov/documents/8-pub207_60743_7.pdf

BP 2: Develop and maintain 1 year annual work plan

Key Performance Indicators

A Key Performance Indicator (KPI)[7] describes a measurable indicator that demonstrates how effectively a company is achieving key business objectives. Organizations use KPIs to evaluate and report their progress and success in reaching key objectives. Similarly, KPIs, both financial and nonfinancial, are an important component of the information needed to demonstrate an observatory's progress towards its stated goals and objectives.

In determining what information KPIs should report, KPI designers should bear in mind the audience who will use the KPIs. This includes reviewers of the observatory's performance, including: funding agencies, external panels, and internal management. It also includes operational staff who can influence and impact the target results outcomes being measured. KPIs provide information necessary for an understanding of the performance of the observatory (PriceWaterhouseCoopers 2007).

KPI design is unique to each observatory, as such a statement describing the intent of the KPI and its relevance to observatory goals and objectives is needed to clarify its purpose. Other design considerations include: Audience and distribution, how many KPIs, grouping of KPIs into scorecard report, level of KPI (e.g., by function, operation or process area), time period reported, referencing similar industry standards, benchmarking to peer groups, and proposed sources of measurement data (PriceWaterhouseCoopers 2007).

There are many types of KPIs to consider [8]:

- Process KPIs - measures the efficiency or productivity of a business process
- Input KPIs - measures assets and resources invested in order to generate results
- Output KPIs - measures the financial and nonfinancial results of activities
- Leading KPIs - measures activities that have a significant effect on future performance. These inform performance of outcome measures, being a predictor of success or failure
- Lagging KPI - measures success or failure after an event, includes most financial measures
- Outcome KPI - measure results in terms of generated benefits, as a quantification of performance
- Qualitative KPI - measures descriptive characteristics, an opinion, a property or trait, includes customer satisfaction
- Quantitative KPI - measures a quantifiable characteristic by counting, adding, averaging numbers. This is most common in performance measurement.

[7] https://www.klipfolio.com/blog/kpi-metric-measure
[8] https://www.slideshare.net/gallasbrows/qualitative-kpi

BP 3: Develop and maintain Key Performance Indicators (KPIs)

Specific, Measurable, Achievable, Relevant, Time-Bound Metrics

A performance metric[9] is a measurement value within a KPI that tracks performance and progress. The difference between KPIs and metrics, simply stated, is that a KPI describes the performance measure and a metric is the number within a KPI that provides a quantifiable value.

Defining KPI metrics that measure and reinforce progress toward the intended objective is difficult. A common consideration when developing performance measures is "what gets measured gets done." A well intentioned metric can have unintentional or unwanted changes to the process or activity be measured.

Useful metrics have two qualities, (1) they are persistent, showing that the outcome of an action at one time will be similar to the outcome of the same action at a later time; and (2) they are predictive, demonstrating a causal relationship between the action and the outcome being measured. Metrics development requires understanding linkages between the cause and effect to assess presumed drivers of the objective. The goal is to make the link between objectives and metrics such that employees can control or influence results through their efforts. Finally, metrics must be regularly evaluated to link activities with the governing objective. The drivers of value change over time, and so
must metrics[10].

Even with abundant data and metrics, insights can be exceptionally tough to come by. Insights here are defined as actionable, data-driven findings that identify future value potential. Developing metrics that lead to insights requires people with deep domain knowledge. Even then, metrics that lead to insights are hard to build[11].

An industry best practice is to develop Specific, Measurable, Achievable, Relevant, Time-Bound (SMART) metrics[12]:

- Specific: Specific metrics are clear and well-defined. Both the grantee and the grantor know what is expected, and the grantor can monitor and assess actual performance against the metrics.
- Measurable: Progress toward metrics is monitored while work is underway. A measurable metric, tracked by the nonprofit, shows when work has been done and when a metric is achieved.

[9] https://www.klipfolio.com/blog/kpi-metric-measure
[10] https://hbr.org/2012/10/how-to-pick-the-right-metrics
[11] https://hbr.org/2012/09/metrics-are-easy-insights-are-hard
[12] https://www.cisco.com/c/en/us/about/csr/community/nonprofits/smart-metrics-for-nonprofits.html

- Achievable: Achievable metrics ensure that everything is in place to meet the metric. If the grantee does not reach its goals, it needs to be able to explain why.
- Realistic: Metrics should be realistic. A metric may have a dependency, such as particular skills, access to resources (computers, tools, etc.), or access to key people and management support. Realistic metrics take these dependencies into account.
- Timely: Descriptions of metrics should include timelines, showing what is required, when. This may include details of delivery, stating (if relevant) where metrics are to be completed. Giving a timeline adds an appropriate sense of urgency and ensures that the metrics do not extend over an unreasonably long period.

BP 4: Develop and maintain Performance Metrics that are Specific, Measurable, Achievable, Relevant, Time-Bound

Performance Metrics Scorecard

Performance scorecards provide a dashboard of key performance indicator (KPI) topics and their associated performance metrics. The scorecard's objective is to quickly convey performance targets and actual results to enable effective assessments of current performance, and identify areas requiring management attention. Multiple scorecards may be needed depending on the complexity of the KPIs.

A performance scorecard primarily refers to a performance management report used by a management team to evaluate and manage the implementation of a strategy or operational activities (Kaplan and Norton 1992). It is a set of performance metrics that provides a broad but comprehensive view of the business objectives and associated status, with the assumption that no single metric can provide a clear indication of overall performance target success.

An individual performance metric doesn't mean a lot. It is how all the metrics work together to measure how well you are addressing your objectives. The concept of a performance metric scorecard weaves together these metrics to provide a single look at the performance of an organization.

Defining performance targets and measuring what you accomplish helps you describe what you do for a variety of audiences including: NSF, panelists reviewing your renewal, the public, congress, and the Office of Management and Budget[13]

Performance metrics scorecards convey performance results from a predefined prior time period. The scorecard also includes performance trend information over a longer pre-defined period, which provides context since one performance period may be a short-term anomaly.

13 https://science.nrao.edu/science/meetings/2013-nsf-large-facility-operations-workshop/documents/ONeil.Performance%20MeasurementMRF-Facilities.pdf

Effectively communicating performance metric results requires thoughtful scorecard development, to ensure that they quickly highlight areas of progress and areas requiring further attention. Revisiting scorecard design periodically ensures continuing alignment with the observatory's objectives. Appropriate distribution of scorecards requires defining the relevant audience for each type of scorecard.

BP 5: Develop, maintain, and distribute Performance Metrics Scorecard

Data Driven Management Reviews

Data-driven performance reviews are regularly scheduled, structured meetings used by organizational leaders and managers to review and analyze data on progress toward key performance goals and other management-improvement priorities. They are used to target areas where leaders want to achieve performance improvements. In addition, reviews should also encourage open dialog which seeks to understand the less clear, qualitative signals and emerging trends. Strategic annual reviews and quarterly goal / priority review meetings are typical meeting frequencies. There is a growing body of evidence that well-run data driven management reviews lead to improved organizational performance (CDC 2017).

From a management perspective, making decisions based on data is critical. Yet it is often difficult to adopt a data-informed culture. In every organization, there are teams and employees who embrace this transition, and those who undermine it. To convert your biggest data skeptics, the first step is to understand the psychology of their resistance. For more information on creating a data-informed culture please see the following resources:

- https://hbr.org/2014/05/whos-afraid-of-data-driven-management
- https://hbr.org/2019/02/companies-are-failing-in-their-efforts-to-become-data-driven
- https://hbr.org/2014/05/an-introduction-to-data-driven-decisions-for-managers-who-dont-like-math

Data driven reviews are also reinforced by the GPRA Modernization Act of 2010 (GPRAMA) and related guidance from the Office of Management and Budget (OMB) direct federal agencies to regularly conduct two different types of performance reviews, annual strategic reviews and quarterly performance reviews where agency leaders and managers assess performance information to determine progress toward meeting goals and objectives, and to update their strategic and performance plans.

Annual strategic review practices federal agencies use to conduct effective reviews include:

- Establish a process for conducting strategic reviews.
- Clarify and clearly define measurable outcomes for each strategic objective.
- Review the strategies and other factors that influence outcomes, and determine which are most important.
- Identify and include key stakeholders in the review.

- Identify and assess evidence related to achieving strategic objectives.
- Assess the effectiveness of achieving strategic objectives and identify actions needed to improve their implementation and impact.
- Develop a process to monitor progress on needed actions.

Quarterly performance data-driven reviews are also reinforced by the GPRAMA, which requires that agency leaders, at least once a quarter, must review and analyze data on progress toward their priority goals. These reviews should examine the progress over the most recent quarter, overall trends, the likelihood of meeting the planned level of performance and, if necessary, strategies to improve performance.

Quarterly data-driven performance review practices federal agencies use to conduct effective reviews include[14]:

- Leaders should use data-driven reviews to drive performance improvement.
- Key players should attend reviews to facilitate problem solving.
- Ensure alignment between agency goals, program activities, and resources.
- Hold managers accountable for diagnosing performance problems and identifying strategies for improvement.
- Ensure that the agency has the capacity to collect accurate, useful, and timely performance data.
- Ensure that agency staff has the skills to analyze and clearly communicate complex data for decision making.
- Enable meaningful performance discussions through rigorous preparation.
- Conduct reviews regularly and frequently.
- Ensure that participants engage in rigorous and sustained follow-up on issues identified during reviews.

Common factors negatively impacting management review success include (Behn 2008):

- No clear purpose defined.
- No one has specific responsibilities.
- Review meetings are held irregularly, infrequently or randomly.
- No one person is authorized to run meetings.
- No dedicated analytic staff.
- No follow-up.
- No balance between accountability and support.

Well designed KPIs and performance metrics enable effective data-driven management reviews by focusing on pre-defined measures of success. Management reviews provide an opportunity

[14] https://www.gao.gov/key_issues/data-driven_decision_making/issue_summary

to adjust and improve on performance measures within the context of strategic mission goals and annual objectives.

BP 6: Perform data-driven management reviews using KPIs and performance metrics

Performance Gap Analysis and Continuous Improvement

Many factors influence an organization's ability to effectively achieve performance targets. External factors such as budget constraints, technology improvements, weather related equipment loses, and evolving science research needs create a dynamic environment which impact an organization's ability to achieve performance targets.

Much of observatory science, and big data in particular are evolving quickly. KPIs can quickly become outdated if they miss new methods. In addition, observatory users can have very different tolerances for observatory products which impact performance measurement targets (e.g. data quality metrics about data satisfaction should seek to differentiate users' ratings based on their respective levels of quality demand).

Frequent data driven management reviews using performance scorecards provides a forum to discuss the gaps between performance targets and current actual results. Gap analysis provides management with data driven information to respond with targeted corrective actions.

As discussed earlier, performance targets represent a conceptual measurement of the observatory’s science plan and concept of operations plan progress toward implementation success. The metric can provide a relevant indicator of whether progress is occurring as planned. It does not provide information for the complex factors impacting performance.

Continuous process improvement is required to ensure the stated performance objectives and their associated performance measurements remain true. Performance targets, current results and gap analysis provide information for areas of potential improvement.

Potential improvements include examining the relevance and accuracy of the performance measurements (KPIs and metrics). The annual work plan processes described above provides a mechanism to revisit the entire performance measurement supporting structure, including the observatory’s science plans and concept of operations plans.

Performance management and the organizations internal capabilities and influence of external factors are dynamic, which makes continuous performance measurement hierarchy process improvement necessary.

BP 7: Perform performance gap analysis and continuous process improvement

The Five Traps of Performance Measurement

Performance measurement has many potential pitfalls and traps. The following are five of the most common traps in designing and managing using performance metrics[15].

Trap 1: Measuring Against Yourself

In order to ensure that you are accurately measuring the success of your observatory, it is recommended that your performance metrics be based on industry standards and not simply reflecting the success of an observatory's current endeavors.

Trap 2: Looking Backward

A performance measurement system that uses measures that lead rather than lag, in other words are indicators of future performance, will help inform decisions made now are going to help the observatory in the future.

Trap 3: Putting Your Faith in Numbers

Good or bad, the metrics in your performance measurement system all numbers. The problem is that numbers-driven managers often end up producing reams of low-quality data. Meaningful action is based on thoughtful context driven design and evaluation of metrics.

Trap 4: Gaming Your Metrics

It is impossible to prevent people from gaming numbers, in other words making numbers look good without actual progress. This can occur no matter how high performing the organization. Diversifying and creating associations between metrics will help to prevent gaming as it is harder to game several of metrics at once. Additionally, varying measurement boundaries by defining responsibility more narrowly or by broadening it (individual or team responsibility) will help reduce metric gaming.

Trap 5: Sticking to Your Numbers Too Long

Performance assessment systems seldom evolve as fast as organizational objectives. It is important to be very precise about what needs to be assessed, be explicit about what metrics are assessing it, and make sure that everyone is clear about both.

Performance Metrics Examples

The observatory performance metrics examples provided highlight topical areas where metrics may be developed. Metrics examples are conceptual to generate discussion, it is not the intent

[15] https://hbr.org/2009/10/the-five-traps-of-performance-measurement?referral=03759&cm_vc=rr_item_page.bottom

of these examples define or recommend specific performance metrics. Each organization should develop their own performance metrics based on their observatory, mission, goals, objectives, and key performance indicators. The examples do not include all metric topic areas, for instance examples do not include financial management, budget performance, staff performance or operational efficiency.

Data Product Quality, Data Delivery and User Support Performance Metrics

Observatory performance metric examples for Data Product Quality, Data Delivery, and User Support are interrelated and dependent on each other. For additional context, please refer to the Data Product Quality Best Practices White Paper.

In general, for data availability, an observatory performance metrics best practice would be to plan for and execute performance metrics for data available from the observatory (fine-scale instrument level to the coarser scale). These metrics include public-facing and internal information. Public-facing information allows users to identify performance information necessary to choose the data they are interested in accessing. Internal information informs community outreach and capacity building efforts, monitors observatory operation and maintenance, and monitors the cost-for-effort of each supported instrument type and data product to inform future budgetary decisions (expansion, upgrade, deprecation, etc.) (EUMETSAT 2018).

In general, data product quality metrics are objective, quantitative measures of achieving (or failing) to produce data products at predefined quality standards. For example, completeness (what data are missing or unusable), conformity (what data are stored in a non-standard format), consistency (what data values have conflicting information), accuracy (what data are incorrect or out of date), duplication (what data are repeated), and integrity (what data are missing).

The setting of simple, easy to quantify targets can lead to a rapid improvement in data quality. A target such as to cut the percentage of new poorly-geocoded records in half every six months for two years can lead to total cut in the error rate of 94% (Redman 2001). Such targets should focus on (Chapman 2005):

- clear and aggressive time frames.
- rates of improvement rather than actual quality values.
- clear definitions.
- targets that are simple and achievable.

Examples of Data Product Quality and Data Delivery Metrics

- Data Availability/Completeness - % available data, (available data / all possible data)
- Data Product Conformity - % data available in a standard data format

- Data Identification - % of data with persistent identifier
- Data Delivery Latency - time between instrument measurement and availability to public
- Automated QC - % of data successfully processed through automated testing
- Human-in-the-loop QC - % of flagged data processed through manual quality control
- Calibration QC - % of data processed with pre and post calibration data
- Comparator QC - % of data processed with validation/comparator measurement (in situ, satellite)

Examples of User Support Metrics

- Self Service - # of views of online tutorials or recorded webinars
- Self Service - # of online training material downloads
- Advanced Support Service - # of registered users
- Advanced Support Service - # of online chat sessions
- Advanced Support Service - # of user phone calls
- Advanced Support Service - # of online community posts on moderated blog
- Help Ticket Responsiveness - time between ticket generation and closure
- Help Ticket Satisfaction - user satisfaction metric captured at closure
- User Website Analytics - frequency, duration, path taken, # of clicks, areas of interest
- User Satisfaction - user satisfaction metric captured from survey or website input

Metadata Performance Metrics

Metadata performance metrics is an evolving and complex topic. Literature research indicates that metadata standards and file formats are varied, although many community efforts to standardize are ongoing. It is beyond the scope of this white paper to address this topic in depth, however, references to example methods are provided below.

Examples of Metadata Metrics

- Data Provided with Metadata - % of data provided to users with accompanying metadata
- Repository Metadata - % of data within repository with accompanying metadata
- Validated Metadata - % of metadata provided with data that is validated
- Automated Metadata - % of metadata created using automated process
- Community Alignment - % of metadata aligning with community standards

Interoperability Performance Metrics

Literature research indicates interoperability performance metrics is an evolving and complex topic. It is beyond the scope of this white paper to address this topic in depth, however, references to example methods are provided below.

Interestingly, an often cited method of measuring interoperability performance relies on best practices integrated into an interoperability Capability Maturity Model (CMM) (e.g. Kasunic

2004). This is then used in determining the current CMM level of your observatory's interoperability, selecting the aspiration CMM level to attain, and measuring progress toward the next CMM level. See the Data Product Quality Best Practices white paper, Interoperability section.

References to other example methods are included below:

"Measuring Systems Interoperability: Challenges and Opportunities," reviews the state of the practice in interoperability and describes the Levels of Systems Interoperability (LISI) Model (Kasunic 2004). This model, although immature, provides a structured and systematic approach for assessing and measuring interoperability throughout the system life cycle. A summary of recommended measures that could promote systems interoperability in the Department of Defense (DoD) is also presented.

"The Health Information Systems Interoperability Maturity Toolkit" consists of a Health Information Systems (HIS) interoperability maturity model, a maturity assessment tool, and this guide for users of the model and the tool (MEASURE Evaluation Carolina Population Center 2017).

"A Qualitative and Quantitative Approach for Measuring Interoperability" introduces a method for measuring interoperability (Knight et al. 2017).

"How to measure interoperability: Concept and Approach," discusses interoperability performance measurement using concepts, challenges and barriers (Ducq and Chen 2008).

Interoperability performance metrics theortically address the three basic dimensions concerning enterprise interoperability, which can be identified as follows (Ducq and Chen 2008):

- Interoperability concerns define the content (or aspect) of interoperation that may take place at various levels of the enterprise (data, service, process, business)
- Interoperability barriers identify various obstacles to interoperability in three categories (conceptual, technological, and organisational)
- Interoperability approaches represent the different ways in which barriers can be removed (integrated, unified, and federated)

Interoperability performance metrics may consider many dimensions, including technical, informational, organizational. An informative example is provided by the GridWise Architecture Council (GWAC) interoperability framework (see Figure 1) below. It conceptually organizes interoperability topic areas to consider, which provides a foundation for defining and prioritizing categories for interoperability.

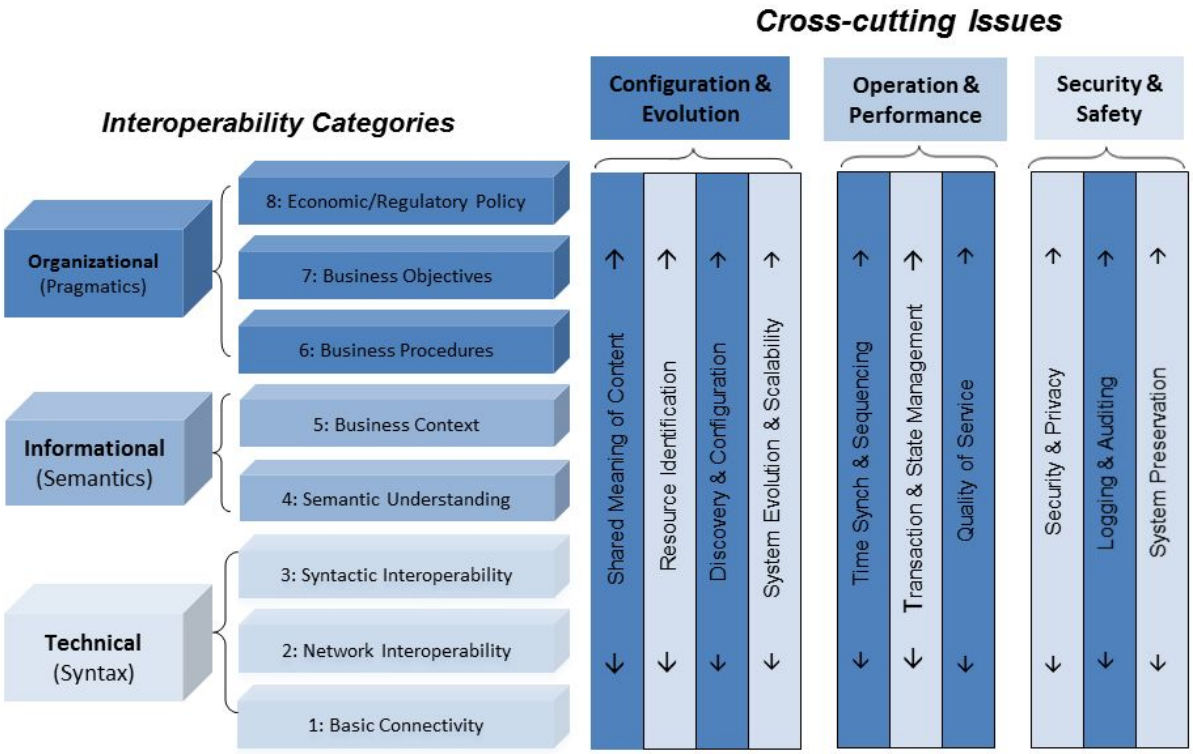

Figure 1. GWAC Interoperability Context-Setting Framework (GridWise Architecture Council. 2008).

Observatory Science Impact and Community Engagement Metrics

Observatory science impact performance metrics measure researcher outcomes and supporting activities including community engagement activities. It should be noted that for the purposes of this paper, community engagement refers to the engagement of other research scientists, it does not refer to engagement of the general public or for educational purposes. Some science impact performance metrics are enabled by data identification and tracking, for example scholarly articles published using observatory data.

While the performance metrics of supporting community engagement activities can provide a picture of how much awareness and engagement has resulted from community engagement activities, they do not necessarily indicate direct science impact outcomes.

For direct science impact outcomes, quantifying actual use and uptake of the data and findings comes through traditional tracking of the number of scientific publications acknowledging the data and their associated citation factors, such as the *h* index (Hirsh 2005). Alternatively, Article-Level Metrics (ALMs) can be used to quantify the reach and impact of published research[16]. ALMs seek to incorporate data from new sources (such as social media mentions)

[16] https://www.elsevier.com/authors/journal-authors/measuring-an-articles-impact

along with traditional measures (such as citations) to present a richer picture of how an individual article is being discussed, shared, and used.These metrics signify when an observatory or findings based on observatory data were used to help further scientific discovery.

The NSF supports the use of both outreach and impact metrics. Per the MREFC Facilities Manual, “metrics and performance goals or targets should include objectives related to educational outreach and broader societal impacts, in addition to research goals of the operating facility” (NSF 2018).

Some of these science impact performance metrics are enabled by data identification and tracking as using a Persistent Identifier, such as a DOI, to search for scholarly articles published using observatory data.

Examples of Science Impact and Community Engagement Metrics:

- Number of observatory workshops and conference presentations by observatory staff
- Number of workshops, trainings and conferences attended by observatory staff
- The number--and growth in number--of participants (new and repeat) in community engagement activities, including geographic location of participants/users
- Number of community members participating in observatory cruises
- User Website Analytics - frequency, duration, path taken, # of clicks, areas of interest
- User Website Analytics - number of new and repeat users (User adoption metrics)
- Size of newsletter distribution list
- Number of social media followers as well as active engagers on online forums
- Number of media releases (news, broadcast, online) distributed
- Number of media placements citing original media release
- Numbers of comments on draft documents posted for community feedback
- Degree of social influence based on social media signals using Klout score (Budden and Michener 2018)
- Number of observatory scholarly (peer reviewed) publications
- Number of observatory scholarly (peer reviewed) publication views, downloads, citations
- Number of unsolicited requests to partner on observatory publications
- Number of proposals prepared and submitted using observatory data/infrastructure
- Number of proposals funded using observatory data/infrastructure

Conclusion

Observatory performance metrics are developed and maintained within the context of an observatory's overall mission, goals, objectives and capabilities in order for observatory managers to measure and improve the success of their observatory in meeting user needs. Generally, for an observatory these are described in the Science Plan, Concept of Operations Plan (ConOps), and Annual Work Plan (AWP).

A typical tool to use for a performance metric is the Key Performance Indicator (KPI) that describes a measurable indicator that demonstrates how effectively a company is achieving key business objectives. These indicators and supporting performance metrics are developed and maintained within the context provided by the Science Plan, ConOps and AWP. Performance scorecards are KPI and metrics reports that demonstrate progress toward observatory goals and objectives. Data-driven management reviews use KPIs, performance metrics and performance scorecards to perform performance gap analysis and inform continuous process improvement.

Not only do funding agencies require these metrics to justify continued funding, but the metrics are useful for the observatory to help define areas of success or areas that require more attention and resources. Additionally, going through the process of defining performance metrics helps organizations define and focus on areas thought to be critical to observatory success. This will allow for more informed allocation of resources. Organizations should be cautious, however, not to be overly fixated on metrics as that can lead to unintended results. It is important to always revisit the objective target of a metric and to adjust metrics to stay on track and take into account improvements in underlying baseline needed.

Best practices described in this white paper are recognized to be challenging to implement. Each observatory has its own priorities and available resources, as such, the best practices described are aspirational. This best practice white paper objective is to provide a simplified, easy to understand and apply guide for self-assessment and planning. It does not represent a guide for technical assessments or implementation.

References

Behn, R.D. 2008. The Seven Bif Errors of PerformanceStat. Harvard University Rappaport Institute for Greater Boston Policy Briefs. Accessed online: https://www.hks.harvard.edu/sites/default/files/centers/taubman/files/performancestat.pdf

Budden A.E. and W.K. Michener. 2018. Communicating and Disseminating Research Findings. In: Recknagel F., Michener W. (eds) Ecological Informatics. Springer, Cham

Center for Disease Control. 2017. Data Driven Reviews: Best practices and Lessons with Agency Priority Goals in Mind. Program Performance and Evaluation Office, Centers for Disease Control and Prevention. https://www.pic.gov/sites/default/files/Data%20Driven%20Reviews%20by%20CDC.pdf

Chapman, A. D. 2005. Principles of Data Quality, Version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

Ducq, Yves & Chen, David. (2008). How to measure interoperability: Concept and Approach. In 14th International Conference on Concurrent Enterprising, June 23–25. Lisbon: A New Wave of Innovation in Collaboration Networks.

EUMETSAT. 2018. Operational Services Specification, v2H. EUM/OPS/SPE/09/0810. 128 pp.

Government Performance and Results Act of 1993. In 103rd Congress. Congressional Record. 1993.

GridWise Architecture Council. 2008. GridWise Interoperability Context-setting Framework, v1.1, 52 pp.

Kamensky, John M. 2011. GPRA Modernization Act of 2010 Explained. IBM Center for the Business of Government http://www.businessofgovernment.org/report/gpra-modernization-act-2010-explained

Kaplan, R. S. and D. P. Norton. 1992. The Balanced Scorecard – Measures That Drive Performance. Harvard Business Review (January–February): 71–79.

Kasunic, M. 2004. Measuring Systems Interoperability: Challenges and Opportunities. Software Engineering Measurement and Analysis Initiative, Carnegie Mellon University. Technical Note: CMU/SEI-2004-TN-003

Knight, M.R., A. Khandekar, B. Nordman, and D. Narang. 2017. A Qualitative and Quantitative Approach for Measuring Interoperability. GRID Modernization Laboratory Consortium. PNNL-26412. Accessed online June 2019: https://gridmod.labworks.org/sites/default/files/resources/InteropIMMTool2017-04-22.pdf

MEASURE Evaluation Carolina Population Center. 2017. Health Information Systems Interoperability Maturity Toolkit: User's Guide. Chapel Hill, NC. Accessed online June 2019: https://www.measureevaluation.org/resources/publications/tl-17-03a

National Science Foundation. 2018. Major Facilities Guide. Draft for Public Comment, NSF 19-XX. National Science Foundation, Arlington, VA. https://www.nsf.gov/bfa/lfo/docs/Major_Facilities_Guide_2019_Draft_For_Public_Comment_December_2018.pdf

PriceWaterhouseCoopers, LLP. 2007. Guide to key performance indicators. Accessed online June 2019: https://www.pwc.com/gx/en/audit-services/corporate-reporting/assets/pdfs/uk_kpi_guide.pdf

Appendix

Best Practice Self-Assessment Tool

The best practice self-assessment tool enables an existing or new organization to assess their current observatory performance metrics capabilities and maturity level. This tool can also be used to identify steps to achieve the next aspirational level. This white paper is intended to provide a Self Assessment Tool for an organization to identify and plan for improvements in people, process, and technology that support observatory performance metrics.

Steps for Using the Self-Assessment Tool

1. Review Best Best Practices List
2. Review Figure 1: Example of a completed best practice self-assessment
3. Determine Self Assessment Capability Scoring
4. Determine Maturity Levels

1. Best Practices List

BP 1: Develop and maintain a Science Plan and ConOps strategic plan
BP 2: Develop and maintain 1 year annual work plan
BP 3: Develop and maintain Key Performance Indicators (KPIs)
BP 4: Develop and maintain Performance Metrics that are Specific, Measurable, Achievable, Relevant, Time-Bound
BP 5: Develop, maintain, and distribute Performance Metrics Scorecard
BP 6: Perform data-driven management reviews using KPIs and performance metrics
BP 7: Perform performance gap analysis and continuous process improvement

2. Example Of Completed Best Practice Self-Assessment

The example below displays one potential combination of capabilities, which results in maturity levels for a hypothetical observatory. Each observatory will have different combinations of capabilities, which aggregate to a certain maturity levels. For example, one observatory may excel at tracking and reporting data citations, whereas another may excel at providing data citation guidance. A simplified capability scoring method is described in the next step.

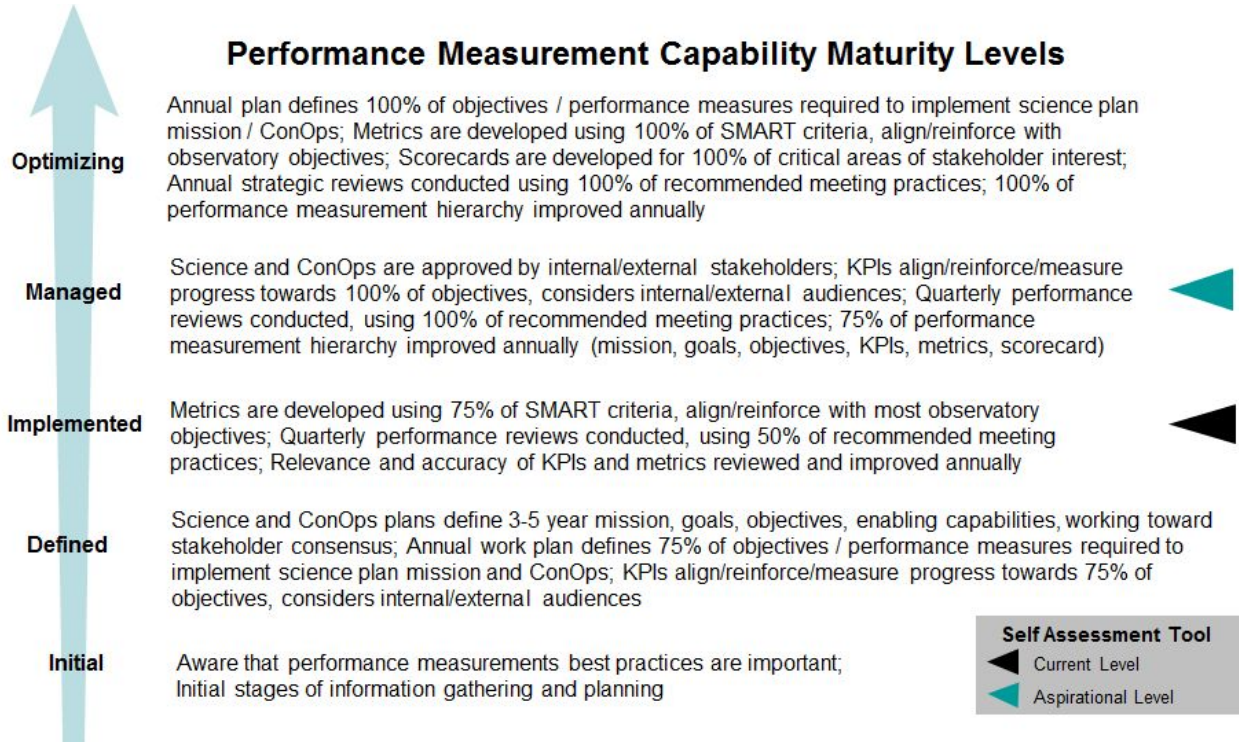

Figure 1: Example of a completed best practice self-assessment

3. Self Assessment Capability Scoring

For each best practice, determine the capability maturity score for your observatory. Only select one capability score per best practice. It is assumed each capability score is inclusive of prior score. Note: Score assumes if capability maturity not present, score is 0.

BP 1: Develop and maintain a Science Plan and ConOps strategic plan

- Plans define 3-5 year mission, goals, objectives, enabling capabilities, working toward stakeholder consensus - 1 point
- Plans define 3-5 year mission, goals, objectives, enabling capabilities, approved by internal/external stakeholders - 2 points

BP 2: Develop and maintain 1 year annual work plan

- Annual plan defines 75% of objectives / performance measures required to implement science plan mission / ConOps - 1 point
- Annual plan defines 100% of objectives / performance measures required to implement science plan mission / ConOps - 2 points

BP 3: Develop and maintain Key Performance Indicators (KPIs)

- KPIs align/reinforce/measure progress towards 75% of objectives, considers internal/external audiences -1 point

- KPIs align/reinforce/measure progress towards 100% of objectives, considers internal/external audiences -2 points

BP 4: Develop and maintain Performance Metrics that are Specific, Measurable, Achievable, Relevant, Time-Bound

- Metrics are developed using 75% of SMART criteria, align/reinforce with most observatory objectives - 1 point
- Metrics are developed using 100% of SMART criteria, align/reinforce with observatory objectives - 2 points

BP 5: Develop, maintain, and distribute Performance Metrics Scorecard

- Scorecards are developed for 50% critical areas of stakeholder interest - 1 point
- Scorecards are developed for 100% of critical areas of stakeholder interest - 2 points

BP 6: Perform data-driven management reviews using KPIs and performance metrics

- Quarterly performance reviews conducted, using 50% of recommended meeting practices - 1 point
- Quarterly performance reviews conducted, using 100% of recommended meeting practices - 2 points
- Annual strategic reviews conducted using 100% of recommended meeting practices - 3 points

BP 7: Perform performance gap analysis and continuous process improvement

- Relevance and accuracy of KPIs and metrics reviewed and improved annually - 1 point
- 75% of performance measurement hierarchy improved annually (misson, goals, objectives, KPIs, metrics, scorecard) - 2 points
- 100% of performance measurement hierarchy improved annually (misson, goals, objectives, KPIs, metrics, scorecard) - 3 points

4. Determine Maturity Levels

Add up your capability score points to determine your current maturity level:

Initial Level	0 points
Defined Level	1-3 points
Implemented Level	4-6 points
Managing Level	7-11 points
Optimizing Level	16+ points

Identify your aspirational maturity level by selecting a desired best practice capability score. Add up your desired capability score points to determine your aspirational maturity level.