RAPTR Design:
How to structure project data

Jim Falter, Environmental Scientist, CNRA

For additional info contact msu@resources.ca.gov

MSU Workshop on Capital and Green Infrastructure - 3 Jun 2021

Today I'd like to talk about one core design element of RAPTR: How it will breakdown and organize project information and translate it into structured, machine-readable data.

How people describe a project...

'The CNRA Bond Program is overseeing the provision of $500,000 in funding from Prop 27 to Parks 'R' Us to develop the Natomas Bike Park through the Improving Bicycle Health Program. This project will provide bicycle owners living in and around Natomas with a fully secure recreational area where their bicycles are safe to socialize as well as roam free and unencumbered; thus, greatly improving the health of both bicycles and their owners. Completion of the project will further provide a direct link between two high traffic bike paths used by residential commuters; thus, facilitating a reduction in street traffic around Sacramento as well as a reduction in net city GHG emissions. The first phase of the project will involve the fee title acquisition of four ~0.5-acre vacant residential lots (156-201-0743, 156-201-0744, 156-205-0613, 156-205-0617) which will then be used to develop the bike park. The second phase of the project will involve the development of various park features including a paved high-traffic bike path connecting two existing bike commuter paths, a fully enclosed special-use area for bikes to move unencumbered, and an open covered community shop equipped with permanently secured bike stands and tools. Development of the project will also involve the planting of 50 native trees as part of a broader climate-change resilient landscaping plan. The total cost of the project will be $1.2 million with additional contributions of $500,000 and $200,000 being made from Sacramento Parks and Friends of Natomas' Bikes; respectively. Sacramento Parks will further be responsible for management of the park (including all O&M) in perpetuity following completion of the project.'

Here is an example of how a person might describe a hypothetical project acquiring some residential property in Natomas and developing it into a bike park. It describes all the sources of money funding the project, the purpose of the project, what land will be purchased, and how the subsequent park will be developed. For many Program Staff trying to review projects like this years after they were completed, it is simply a '*wall of words*'.

How computers describe a project…

FIELD	VALUE	FORMAT
ProjectID	"CNRA-123-4567"	text
ProjectName	"Natomas Bike Park"	text
GranteeName	"Parks 'R' Us"	text
ProgramName	"Improving Bicycle Health Program"	text
AdminOrg	CNRA	system-defined
ProjectDescription	"This project will provide bicycle…"	text
AcqParcels	{156-201-0743, 156-201-0744, 156-205-0613}	text
ProjectArea	2.07 [acres]	number
AcqType	{Fee Title, Fee Title, Fee Title, Fee Title}	system-defined
PropertyManager	"Sacramento Parks"	text
TotalCost	1.2e6 [dollars]	number
FundingAmounts	{5e5, 5e5, 2e5} [dollars]	number
FundingSources	{"Prop 27", "Sacramento Parks", "Friends of Natomas' Bikes"}	text

MSU Workshop on Capital and Green Infrastructure - 3 Jun 2021

Here is how a database would describe the same project by decomposing it and translating it into individual pieces of data assigned to pre-specified data fields. This kind of decomposition of project information will allow us to make more refined queries across thousands of projects based on specific attributes of interest such as…

1. What other State, Federal and non-governmental organizations co-funded the project? And if so, by how much?
2. What deliverables and benefits did the project provide?
3. How many other State projects produced similar deliverables or benefits?

Despite the necessity of the project data translation process, I want to assure future users that RAPTR will nonetheless still require the submission of written narratives describing the scope and activities of a project as defined in the finalized Grant Agreement or Contract. It will also provide staff with a diary that can keep track of communications with project personnel as well as provide internal commentary on project implementation. This will make it easier for Program Staff to document the '*narrative arc*' of project development without having to re-interpret what is happening from the individual machine-readable data.

Defining an Activity and its Benefits

Agent → Action → Asset → Benefit

MSU Workshop on Capital and Green Infrastructure - 3 Jun 2021

In the last slide I gave some off-the-cuff examples of how the data implicit to a project narrative could be reasonably parsed just to demonstrate how the translation of information from narrative prose to machine-readable data could occur. In reality, however, this process actually requires a bit more forethought to ensure that future projects managed by all programs (and soon-to-be 27 departments, conservancies, councils and boards) can be broken down according to a common data structure that is flexible and generic enough to cover the broad scope of assets and activities administered underneath Agency (activities that can range from the construction of hydroelectric dams to educating our communities about the importance of habitat conservation) and yet provide enough granularity to capture all the pertinent details necessary for describing the activities completed and benefits achieved.

One simple way to describe the activities of CNRA programs would be to break them down into four fundamental components: An *Agent* conducting an *Action* on a natural resources *Asset* for the purpose of achieving a *Benefit*. This chain of causality represents the fundamental '*grammar*' with which all project activities across all Agency programs can be translated into a machine-readable data structure.

Simple Examples

	Who		What		Why

Agent	Action	Asset	Benefit
DWR ⇨ Stantec	Construct	Dam	Reliable Water Supply
CDFW ⇨ River Partners	Restore	San Dieguito River	Conserving Habitat
Parks ⇨ City of Shafter	Develop	Shafter Community Park	Access to Recreation
Cal FIRE ⇨ City of Escondido	Educate	Stakeholders	Urban Greening
SCC ⇨ MarinLink	Train	Volunteers	Conserving Habitat
OPC ⇨ UC System	Monitor	Coastal Waters	Improving Water Quality
CTC ⇨ CTC	Regulate	Upper Truckee Marsh	Conserving Species
SNC ⇨ Sierra Institute	Improve	Wood Products Facility	Climate Change Mitigation
WCB ⇨ Tuolumne River Trust	Improve	Roads	Access to Nature

MSU Workshop on Capital and Green Infrastructure - 3 Jun 2021

Here are some simple and less abstract examples of what I mean:

1. DWR, acting through a contract with an engineering firm, implements the construction of a new dam to improve the reliability of the State's water supply.
2. Fisheries and Wildlife, acting through a grant to an NGO to restore the San Dieguito River in the interest of conserving important riverine habitat.
3. State Parks, acting through a grant to the City of Shafter, develops the Shafter Community New Park to provide enhanced access to recreation.

In this sense, it might be easier to think of the Agent as being the '*Who*', the Action and Asset as being the '*What*', and the Benefit as being the '*Why*'. I've added a few other examples to help illustrate this concept.

Types of natural resources Assets

Water Resources	Park Infrastructure	Habitats	Wildlife
Levee	Play Structure	Forest	Chinook Salmon
Detention Basin	Water Feature	Wetlands (Estuarine)	Valley Oak
Pumping Station	Soccer Field	Upland	Delta Smelt
Revetment	Campground	Stream	Spotted Owl
Water Treatment Plant	Parking Area	Riparian	Giant Sequoia
Weir	Interpretive Display	Beach	Nutria (*invasive*)

Of the four project data types just discussed, natural resources *Assets* are perhaps the most diverse and varied when examined over the full range of Agency programs. Here are a few examples of Asset types that could be the objects of various conservation, infrastructure, and park development activities.

Activities defining a Project
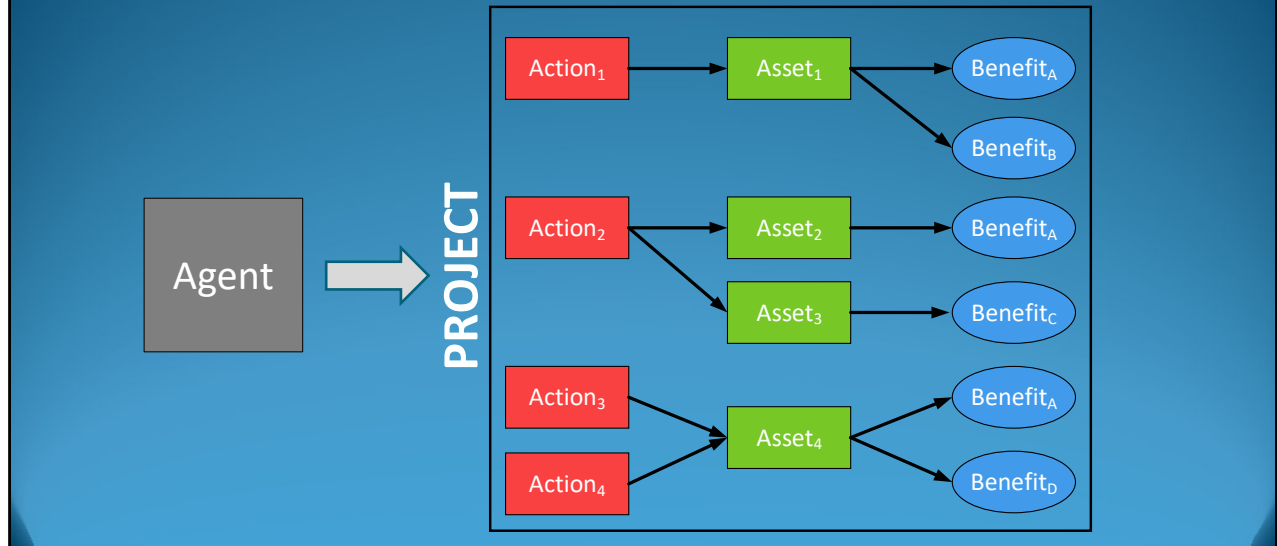
Some Agency-supported projects may be as simple as the examples I've just shown: one agent conducting one action on one asset to yield one benefit. However, project activities can also consist of a single action operating on multiple assets, or many actions operating on a single asset which, in each case, yield one or more benefits. Even further, a single project can involve a variety of different activities.

For example, a hypothetical project could include land being purchased so that a flood detention basin can be constructed around a creek whose banks are concurrently de-armored and restored with native vegetation. To make best use of the area (since it will be flooded for a small fraction of the time), the bottom of the basin is converted into soccer fields and a walking path for local communities to use. The resulting project would thus consist of a multi-disciplinary effort involving 1) a fee title acquisition of the necessary land, 2) the construction of the flood detention basin, dam and overflow, 3) the restoration of stream and riparian habitat present, and 4) the development of additional park infrastructure. This 'one' project would then achieve the multiple benefits of Improved Flood Protection, Conserving Important Habitats, and Enhancing Access to Recreation. The ability of RAPTR to accommodate the translation of a wide range of complex projects was a key issue posed by a number of stakeholders, and this fundamental data structure is how we plan to address that concern.

Validating Project Activities

Project Metrics | Deliverables

Agent | Action | Asset | Benefit
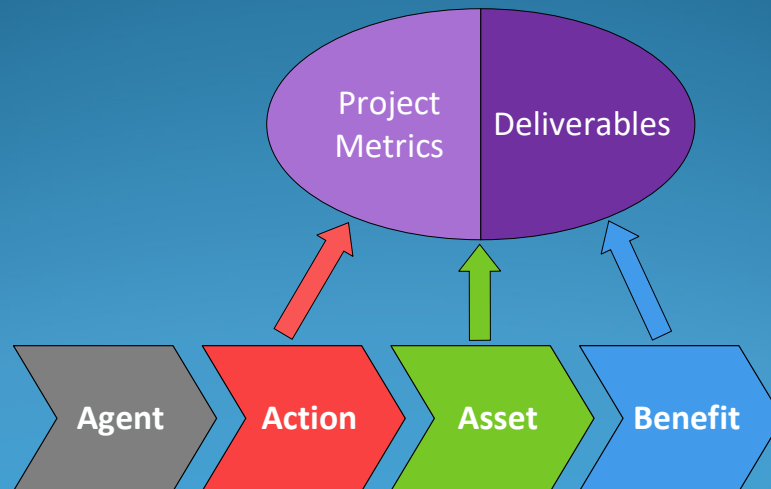
Once the activities of a project have been parsed and classified, then the next step is to identify expected *Deliverables* and *Project Metrics* defining what the project has achieved and validate what was promised in the signed Grant Agreement or Contract. The specific Metrics and Deliverables relevant to a given set of project activities will, of course, depend on the activities defined in the projects; that is the specific Actions, Assets and expected Benefits. One distinction I want to make here is that while from the perspective of Program Staff a '*Deliverable*' might mean some physical object like a pumping station, from the perspective of RAPTR it will mean the documentation validating the existence of the deliverable like the source schematics for the pumping station as well as some kind of final inspection report proving it met the design specifications articulated in the corresponding grant agreement or contract.

Validation Matching is Heterogeneous
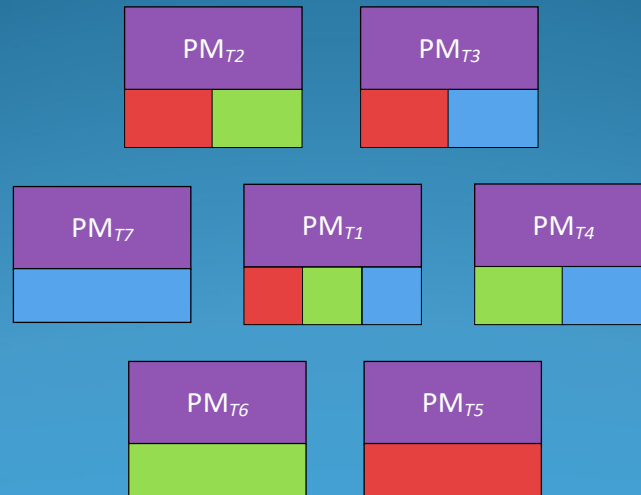
MSU Workshop on Capital and Green Infrastructure - 3 Jun 2021

One slight twist to determining the right set of project metrics and deliverables is that they can depend on different combinations of the project activity data types. For instance, the benefit of Climate Change Mitigation is generally measured by the annual amount of greenhouse gas emissions avoided or sequestered regardless of the action and asset in question. The scope of work conducted on pipelines, levees, and canals is generally measured in linear distance regardless of the action or benefit achieved. Activities involving education and community engagement can be measured in the number of community groups and people engaged regardless of the subject matter over which the engagement is occurring, and the benefits of interest involved.

Assigning Project Metrics & Deliverables

PROJECT

Action$_1$ → Asset$_1$ → Benefit$_A$
Asset$_1$ → Benefit$_B$

Action$_2$ → Asset$_2$ → Benefit$_A$
Asset$_2$ → Asset$_3$ → Benefit$_C$

Action$_3$ → Asset$_4$ → Benefit$_A$
Action$_4$ → Asset$_4$ → Benefit$_D$

$$\sum_{j=1}^{N_A}\sum_{i=1}^{N_{PM}} PM_i \begin{pmatrix} a_j \subseteq \mathbf{A}_i \\ s_j \subseteq \mathbf{S}_i \\ b_j \subseteq \mathbf{B}_i \end{pmatrix}$$

$$\sum_{j=1}^{N_A}\sum_{i=1}^{N_D} D_i \begin{pmatrix} a_j \subseteq \mathbf{A}_i \\ s_j \subseteq \mathbf{S}_i \\ b_j \subseteq \mathbf{B}_i \end{pmatrix}$$
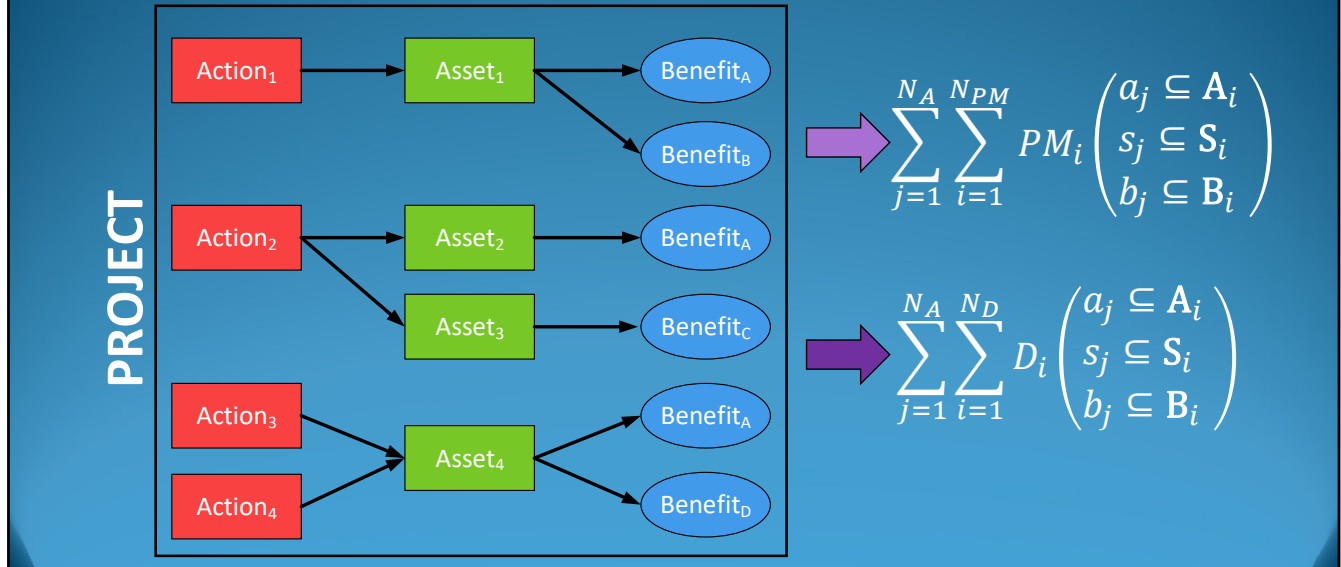
How do we bring these processes together – that is both classify project activities and identify the relevant project metrics and deliverables? Given the heterogeneous nature of the likely matching, it seems that project metrics and deliverables will match a given project activity according to first finding all the project metrics and deliverables relevant to a given activity as defined by the relevant combination of Action, Asset and Benefit and then repeating that cycle for each proposed project activity.

This whole process of data classification and validation assignment may seem a bit tedious and time-consuming compared to the more conventional approach of copying and pasting sections of text from a Project Description, but this is *exactly* the kind of task that data management systems are well-suited to automate and perform on our behalf.

RAPTR will help Users with this...

PROJECT

Action$_1$ → Asset$_1$ → Benefit$_A$, Benefit$_B$

Action$_2$ → Asset$_2$ → Benefit$_A$
Action$_2$ → Asset$_3$ → Benefit$_C$

Action$_3$, Action$_4$ → Asset$_4$ → Benefit$_A$, Benefit$_D$

$$\sum_{j=1}^{N_A}\sum_{i=1}^{N_{PM}} PM_i \begin{pmatrix} a_j \subseteq \mathbf{A}_i \\ s_j \subseteq \mathbf{S}_i \\ b_j \subseteq \mathbf{B}_i \end{pmatrix}$$

$$\sum_{j=1}^{N_A}\sum_{i=1}^{N_D} D_i \begin{pmatrix} a_j \subseteq \mathbf{A}_i \\ s_j \subseteq \mathbf{S}_i \\ b_j \subseteq \mathbf{B}_i \end{pmatrix}$$
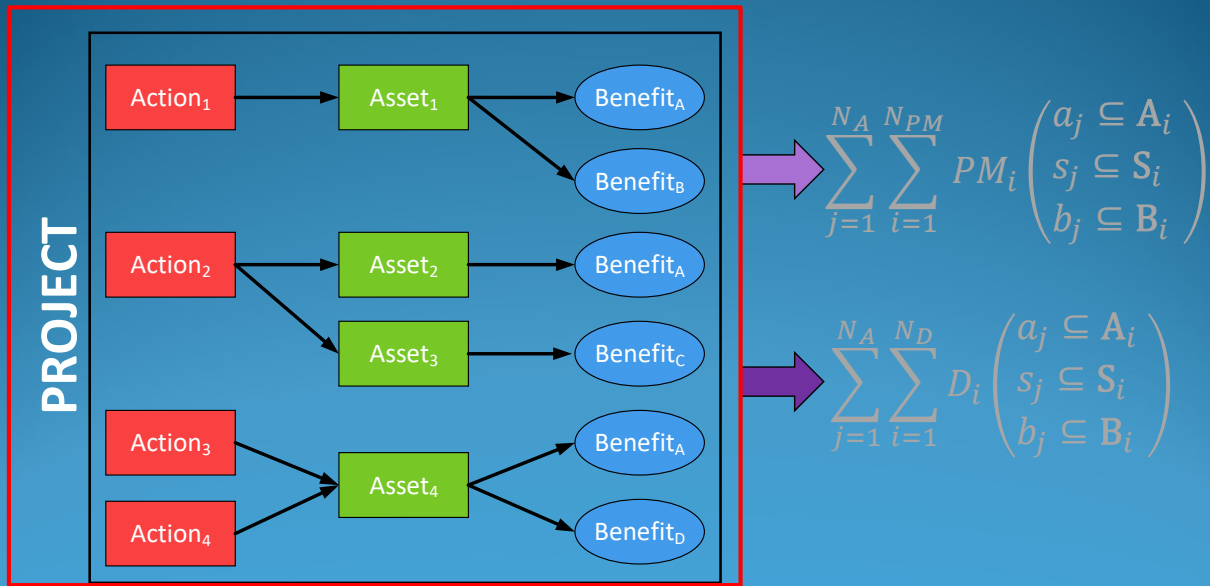
For instance, we are currently pursuing the development of an online tool that will help guide Program Staff in the parsing and classification their project data so that they only have to choose from a limited scope of options relevant to the activities of their project and program. Regardless of how this tool is developed, Program Staff will still have the final say over how they chose to define the Actions and Assets involved in their project as well as the Benefits they expect that project to achieve.

If you can imagine project data types as being similar to baseball cards representing different players from different teams, RAPTR will not tell Program Staff which 'cards' to use in describing their projects, only that they chose those cards from a standardized deck. More importantly, this 'deck' will have been created largely based on input provided by Program Staff and other key stakeholders.

As Program Staff contribute their own suggestions to the library of project metrics and deliverables based on relevant activities aggregated across all programs under Agency, RAPTR will rely on this '*home-grown*' data set to automatically provide a limited but comprehensive subset of potentially relevant project metrics and deliverables from which Program Staff can then chose to apply at their own discretion.

To sum up: The translation of project information into machine-readable data structures will allow for far more rapid and complex assessments of Agency activities and outcomes across all its departments, programs, and corresponding geographic regions of interest. This effort will further obviate the need for Program Staff to conduct the kind of labor-intensive '*fire drills*' needed to answer a single policy or administrative question that, in turn, require the expenditure of hundreds of person-hours locating and reviewing a scattered distribution of individually organized documents years after the projects were completed. We hope this approach will not only help ensure the accuracy and completeness of project data but, at the same time, help Program Staff better track the short- and long-term benefits of the project over time.