Inkan data management "good practices" and applicability to the Federal Data Strategy by Jim Watts

While many in-person conferences this past year have been cancelled due to Covid 19, thanks to improved distance learning technology we have been able to continue to learn. Last Summer I was able to virtually attend Howard Dresner's annual Real Business Intelligence Conference[1]. Normally this would be held on-site at MIT; however, Real Business Intelligence 2020 was a virtual event. All presenters shared interesting facts and observations and one seemed particularly relevant to current-day data management. Gary Urton shared a presentation about knots and beads and Inkan (also spelled Incan) Empire Accounting, based upon his publication *Inka History in Knots*.[2] While the Inkas may not have had the benefit of current-day data management technology with computers and data centers, they were able to create and sustain enterprise data management over time through manual processes. Several of their manual processes seem remarkably similar to concepts we use today. With this in mind there may be applicability from their solutions using manual processes to apply towards current day enterprise data management.

The goal of the Federal Data Strategy, established through the President's Management Agenda in March 2018 is to "Leverage data as a strategic asset". The United States Government's Federal Data Strategy[3] is currently working through an Action Plan to help agencies develop a mature data asset management environment. Currently there are numerous data models across the government. Software vendors, both inside and outside the Federal Government have created their own standards based upon use cases. The mission of the Federal Data Strategy is to leverage the full value of Federal data for mission, service, and the public good by guiding the Federal Government in practicing ethical governance, conscious design, and a learning culture. The Federal Data Strategy includes 10 principles, 40 practices, and 20 action steps.

The purpose of this article is to share several high-level comparisons between Inkan data management standards and current-day enterprise data management. Where possible when discussing specific Inkan groupings we will include related alignment to the Federal Data Strategy. We will not look at all aspects of Inkan data management or all aspects of the Federal Data Strategy, but will focus on the following areas: 1) data and data groupings, 2) administration and training, 3) storage, and finally 4) storytelling and evidence-based policymaking.

[1] https://realbusinessintelligence.com/

[2] *Inka History in Knots: Reading Khipus as Primary Sources*. Austin, TX: University of Texas Press, 2017.

[3] https://strategy.data.gov/

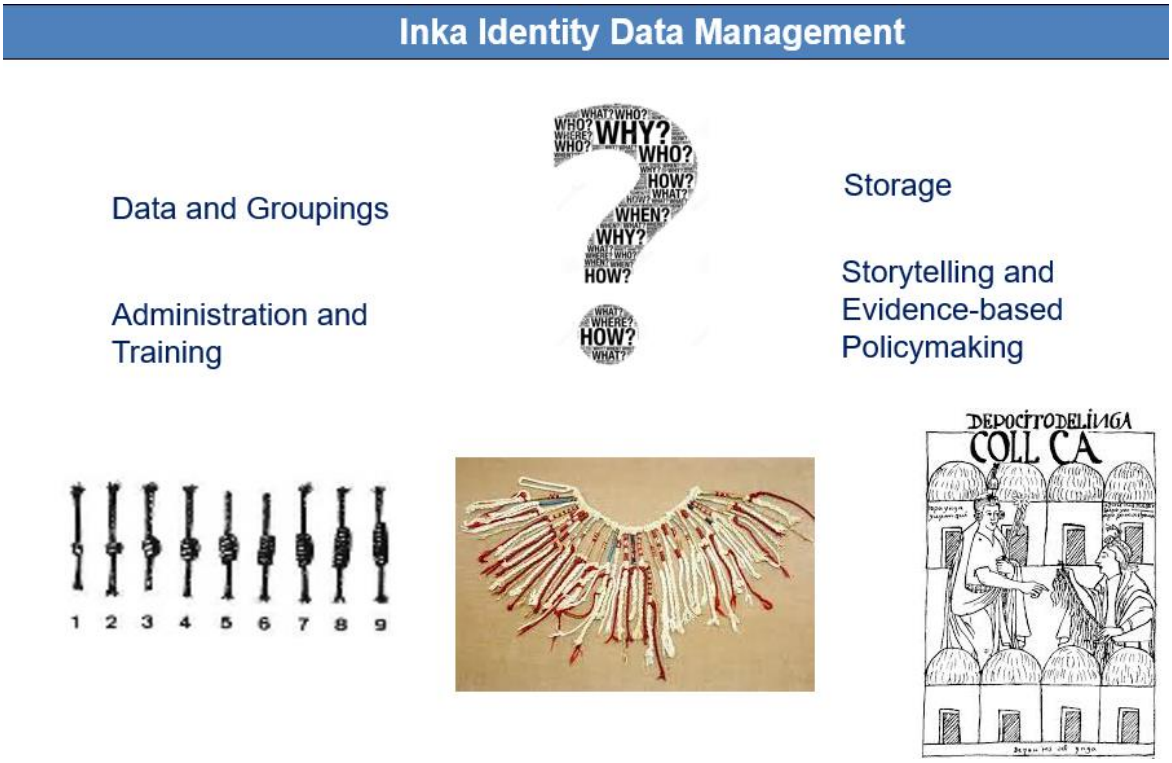

Figure 1: Khipus knot counting with inventories and storage

The Inkan civilization and form of government existed between 1438 and 1533 and during this time the population reached approximately 10 million people. Their civilization expanded through battles, one province at a time. During this time their culture consisted of many of the same concepts we have today, including society and government and an economy. While they didn't have the benefit of modern technology they were able to create a variety of manual processes. In order to help them manage their economy they developed physical measures, numerical counts, mathematics, and accounting using a combination of knots and beads.

Data and Data Groupings

The Inkas needed a way to track numerical information, organized by identities. Identities included names of people, places, things, events, etc. Many of these identities were based upon color coding. For example, the Inkas called a grouping of knots Khipus. The following quote is attributed to Garcilaso de la Vega "When the Inca had conquered a province he had a record made on his knots and beads of the pasture lands, high and low hills, plough lands, estates, mines, salt works, springs, lakes, and rivers, cotton fields, and wild fruit-trees, and flocks of both kinds [llamas and alpacas], including those that produced wool, and those that did not. All these things and many others he had counted, measured, and recorded under separate headings, firstly the totals for the whole province, and then those for each village, and each inhabitant. They measured the length and breadth of the arable land, the cultivable area, and the pasture land. When all the details were known, a full report was made of the whole province."

This breakout of data by groupings is not unlike what we currently track within current-day enterprise data management. The Inkas may have had more success because they were able to prioritize and standardize data and key data sets. The Inkas had an advantage of a smaller population and governance structure, whereas the United States has a much larger population and governance structure.

The Federal Data Strategy consist of twenty action steps, including prioritizing data and data sets. There are three formal action steps (items Action 5, Action 6, and Action 8) within the Federal Data Strategy that help support agency prioritizations:

Action 5: Identify Priority Data Assets for Agency Open Data Plans

Federal Data Strategy	Inka	Discussion
Action 5: Identify Priority Data Assets for Agency Open Data Plans	**Data and Data Groupings** information, organized by identities: people, places, things, events, etc.	List any data sets that your office created, uses, needs, or plans to create "Produced" or "Consumed"

By December 2020, all agencies will identify their initial list of priority data assets for agency open data plans. This list will describe data assets that the agency considers especially valuable to the public interest and therefore intends to make available as open government data. Agency open data plans will include (1) processes and procedures that require data collection mechanisms created on or after January 13, 2019, facilitate open formats, and facilitate collaboration with non-Government entities (including businesses), researchers, and the public for the purpose of understanding how data users value and use government data; (2) identify methods to collect and analyze digital information on data asset usage by users within and outside of the agency, including designating a point of contact within the agency to assist the public and to respond to quality issues, usability issues, recommendations for improvements, and complaints about adherence to open data requirements within a reasonable period of time; (3) develop a process to evaluate and improve the timeliness, completeness, consistency, accuracy, usefulness, and availability of open Government data assets; and (4) includes requirements for meeting the goals of the agency open data plan, including the acquisition.

Action 6: Publish and Update Data Inventories

Federal Data Strategy	Inka	Discussion
Action 6: Publish and Update Data Inventories	**Data and Data Groupings** prioritize and standardize data and key data sets.	Criteria for data sets are those valuable to the public interest and therefore intends to make available as open government data on data.gov

All agencies will update existing comprehensive data inventories in accordance with updated standard metadata that facilitates ingestion by data.gov and search engine optimization. At least every 90 days, all agencies will identify missing or incomplete listings in their comprehensive data inventories and ensure metadata is comprehensive for priority data assets, such as data assets included in agency Open Data Plans, identified in Community of Practice Actions 8, 9, and 10, or as identified by the Chief Data Officer Council (CDO Council).

Action 8: Improve Data and Model Resources for AI Research and Development

Federal Data Strategy	Inka	Discussion
Action 8: Improve Data and Model Resources for AI Research and Development	*Data and Data Groupings prioritize and standardize data and key data sets. *Does not apply	Enhance access to high-quality and fully traceable federal data, models, and computing resources

"Enhance access to high-quality and fully traceable federal data, models, and computing resources to increase the value of such resources for AI R&D, while maintaining safety, security, privacy, and confidentiality protections consistent with applicable laws and policies."

Similar to Inkan Khipus, current-day enterprise data management requires a great deal of planning and grouping. Additional complexity within current-day efforts is based primarily on the sheer volume of sources, data elements and disparity during data creation. One way around this for the federal government is through the adoption of the Federal Enterprise Architecture's Enterprise Conceptual Data Model and Definitions.[4] This will provide a standard set of data groupings and data elements as well as common set of information over time.

Administration and Training

The Inkas relied a great deal upon the trustworthiness of public notaries, called quipucamayos, to measure, count and record each knot and bead. They had to be "believed absolutely" in order for the system of knots and beads to succeed. Data integrity and focus on ethics was built throughout the Inkas process thanks to these quipucamayos. In order to complete their activities, they used the same standards for measures as well as the same standards for colored knots.

The Federal Data Strategy includes two related action steps to prioritize skills and staffing:

Action 4: Identify Opportunities to Increase Staff Data Skills

[4] https://resources.data.gov/standards/

Federal Data Strategy	Inka	Discussion
Action 4: Identify Opportunities to Increase Staff Data Skills	• Public notaries, called quipucamayos, to measure, count and record each knot and bead	• Perform a capacity assessment to assess the coverage, quality, methods, effectiveness, of current staff data literacy and data skills

By June 2020, all CFO Act agencies will have performed a capacity assessment to assess the coverage, quality, methods, effectiveness, of current staff data literacy and data skills. These assessments may be further leveraged by agencies to identify critical data skills, assess their staffs' capacities for those skills, identify any gaps, and take actions to ensure that their Federal workforces are well-prepared to support evidence-building activities pursuant to M-19-23 and the Evidence Act.

The staff data skills assessment should align with, and be a part of, ongoing human capital management efforts designed to ensure agencies have skills and competencies needed to effectively accomplish agency mission. This alignment is important since achieving parity between an agency's data skill needs and its workforce capacity requires repeating this process over time. The assessment has four major components:

1. Identify critical data skills needed for the agency needed to support management of priority data sets
2. Assess the current staff capacity for those data skills to ensure priority data sets can be managed over time
3. Perform a data skills gap analysis to prioritize the agency's needs to identify risks
4. Identify and execute approaches to fill those needs to be able to support management of priority data sets

Action 14: Develop a Data Ethics Framework

Federal Data Strategy	Inka	Discussion
Action 14: Develop a Data Ethics Framework	• Quipucamayos had to be "believed absolutely"	• GSA will develop a data ethics framework

"GSA will develop a data ethics framework to help agencies systematically identify and assess the potential benefits and risks associated with the data they acquire, manage, and use. This framework is intended to help agency staff, managers, and leaders make considered data

acquisition, management, and use decisions to address ethical issues they may encounter throughout the data lifecycle. The framework will be broad and flexible – outlining and describing high-level principles that can be applied to specific circumstances, such as preparing data for the responsible and ethical use of evolving technologies, including mitigating bias when developing AI and machine learning methods and systems.

Once the completed framework and associated implementation tools are published, they will be promoted government-wide for agency use via the CDO Council, ICSP, and various other channels. Fully integrating a data ethics perspective into all aspects of agencies' data management efforts will require substantial and long-term cultural change. It would involve staff at all levels undergoing training to support and refresh data literacy skills and reinforce protocols related to data privacy, confidentiality, and the ethical collection, use, storage, and dissemination of data. Leadership support, spearheaded by CDOs and Statistical Officials, is crucial, as is agency input into the design of tailored ethics frameworks and their implementation.

The identification of critical data skills will naturally be informed by the determination of agency priority questions, including those gathered during the development of the learning agenda (see Action 1) and will be addressed in the agency maturity assessment (Action 3). Thoughtful consideration should also be given to identifying a minimum level of data literacy among all staff, including for those performing roles not traditionally considered data related. The agency should assess the data skills possessed by the current workforce and seek to understand data literacy rates among its staff. The gap analysis should determine how much more of each critical skill is required among staff and facilitate a prioritization of needed skills relative to available resources. Options for increasing staff skills capacity may include new analysis or other software tools, easy-to-use dashboards, additional training and educational opportunities, on-the-job rotational learning experiences, participation in data-related communities of practice, and introducing hiring and retention strategies to address gaps.

Storage

The Inkas constructed buildings called Qolqawasi that served as central storehouses for khipus. Historical khipus were stored along with current khipus. Gary Urton and Alejandro Chu published work in the Latin American Antiquity[5], however, provided further proof how khipus were used in Inka storage facilities.

In 2013 and 2014, Chu was conducting excavations at the site of Inkawasi, located along the southern coast of Peru. The site is located along the famed Inka road system and included large storage facilities built by the empire to collect goods generated by the mit'a labor system. While excavating one of these storage facilities, Chu and his team recovered 29 khipus, four of which were found in association with preserved caches of foods including chili peppers, peanuts, and black beans. This offered an opportunity to compare the accounting records of each khipu with the goods they were presumably used to track.

[5] Article: Urton and Chu [December 2015] Latin American Antiquity Vol. 26, No 4; Cambridge University Press.

In analyzing the patterns of strings and knots found on these four khipus, the authors of the new study note an intriguing pattern. It appeared each khipu included a count of total goods along with a small "fixed count" that appeared to refer to a small percentage of goods that were being set aside. In three of the four cases, although the base numbers were different, the amount set aside turned out to be approximately 2% of the original total count.

While there is some room for debate, Urton and Chu argue in their paper that the most likely interpretation of the numbers laid out by these khipus is as a tax used to support the running of the storehouse. They note that "Inkawasi was a new kind of facility in the Inka imperial infrastructure, and as such, the management of the site stimulated new accounting procedures." The Inka empire was a dynamic entity, growing, developing, and changing all the way up until its downfall at the hands of European epidemic diseases just a few generations later.

The current-day data centers provide similar storage, not for khipus and food, but for computers and data. Data centers serve as storage for computers, applications, and data, not khipus and food. The cost model for data centers has changed from on-premises cost models. As government agencies migrate data to data centers, the cost model may change from a centralized IT funded function to office or program. The Federal Data Strategy includes one general action step with focus on data standards and resources:

Action 11: Develop a Repository of Federal Enterprise Data Resources

Federal Data Strategy	Inka	Discussion
Action 11: Develop a Repository of Federal Enterprise Data Resources	• Buildings called Qolqawasi that served as central storehouses for khipus • Historical khipus were stored along with current	• Cloud and Data.gov • Improved Evidence-based Policy Making

The General Services Administration (GSA), the Office of Government Information Services of the National Archives, and OMB collaborated with stakeholders to develop the website resources.data.gov, a government-wide repository of policies, standards, tools, best practices, and case studies that is required under the Evidence Act.[12] This repository, titled *Federal Enterprise Data Resources,* has been established to provide centralized access to resources related to Federal data management and use in support of agencies as they seek to execute both the FDS and the requirements of the Evidence Act.[13]

Similar in concept to the Inkas, the repository will not store various material goods and khipus, but will store a variety of critical digital resources including:

- **Case Studies**: stories, good practices, and examples of what agencies are currently implementing, such as providing a business case to agencies for open data;
- Detailed data and data types **Data.gov Schema**: The standard metadata schema required for inclusion in data.gov, the Federal data catalog[14]

- Federal Data Strategy: Guiding principles and best practices and associated implementation tools;

- **Playbooks**: Decision points, checklists, and questions that help organizations navigate options,

- **Requirements**: Definitions, laws, policies, and regulations;

- **Skill Builders**: Professional development resources and curriculums to improve data-related skills,

- **Standards**: Guidance on preferred technical formats, licensing, and resources related to implementing data standards for data sharing and

- **Tools**: Tools available to agencies to help execute data goals and begin to share and reuse data across the federal government.

The Federal Enterprise Data Resource will provide a one-stop-shop to help agencies build standard practices with the intent to make agency data accessible as agencies develop statistical evidence to support policymaking while providing transparency and insights for an open government. Continuous access over time will provide the means to update baseline documents while building for the future.

Storytelling and Policymaking

The Inkas used the Khipus to help tell stories. According to text attributed to Martin de Murua (1591) "By these knots, they kept track of the passage of time and when each Inca reigned, the children he had, if he was good or bad, valiant or cowardly, with whom he was married, what lands he conquered, the buildings he built, the service and riches he possessed, how many years he lived, where he died, and what he was fond of."

In many ways the Inkas were more easily able to tell these stories because they had one set of data standards. They recorded data using knots and the knots were grouped into Khipus. The Khipus serve as a record of time.

Current-day systems and data centers were not built with the end goal of storytelling for one person, one place, one thing, one event, etc. The Federal Data Strategy will help align data and data sets to provide evidence and policymaking. The Federal Enterprise Architecture (FEAF) 2.0 Data Reference Model (DRM) provides one approach to dive even deeper into current-day storytelling as well as evidence-based policymaking. As shown below, data reference model can be used to create a visualization, an Enterprise Conceptual Data Model (ECDM) as an authoritative data standard. Here is an extract from the ECDM, represented not in colored knots and beads, but in groupings more appropriate for data management:

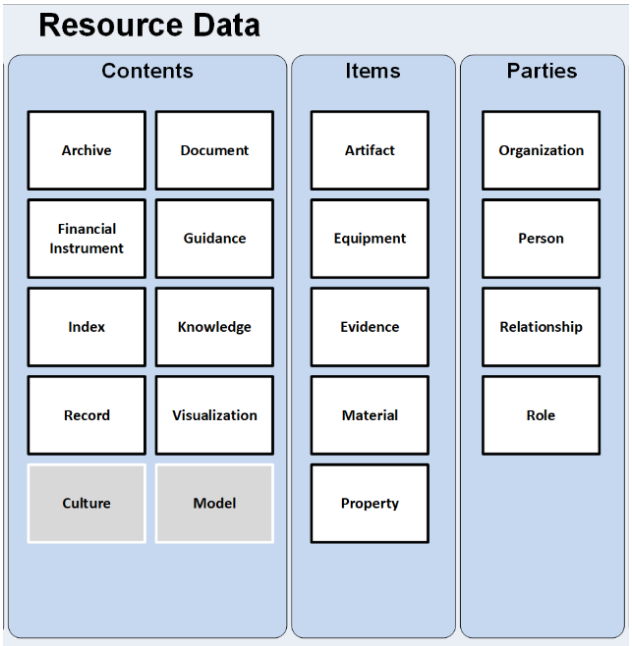

Figure 1: Portion of 100% generic visualization of the Federal Enterprise Architecture (FEAF) 2.0 Data Reference Model (DRM)

Summary

Current-day systems and data centers are far more advanced for reporting and analysis than Inkan khipus and Qolqawas. Current-day systems and data centers are connected to the end user. Application developers continue to create innovative ways to improve. Data access is much faster than Inkan manual methods.

A continuing challenge is that our current-day systems and data centers were created for the most part without common data standards. They are, to some extent, missing the colored knots and beads that helped build the Inkan baseline for data management. The current-day systems were to some degree built as silos conceived by individual developers. As a result, there is much more complexity, complexity around data, privacy, security, ownership, combined data across data centers, and access. As a consequence, it may be more difficult to tell the story of one person, one place, one thing, one event, etc.

Data exchanges need to align with data standards to bridge applications across data centers. The Federal Data Strategy, although ambitious, is vital to current and future evidence-based policymaking and storytelling through data access and enterprise data management. The Federal Data Strategy and data standards will help bridge current gaps. Let's don't forget the Inkas. Even though a pandemic wiped out their civilization, there may be more to learn from the Inkan enterprise data management, ethical governance, conscious design and learning culture.

Article: Urton and Chu [December 2015] Latin American Antiquity Vol. 26, No 4; Cambridge University Press.

Book: Urton, [2017] *Inka History in Knots: Reading Khipus as Primary Sources*. Austin, TX: University of Texas Press, no specific pages cited.

ABSTRACT

Enterprise data management remains a challenge for many companies and government agencies. Good practices can help share insights to help companies and government agencies improve performance of their data. Good practices can include manual techniques as well as automation. One technique the Ancient *Inkas* used to create manual standards was through the use of a combination of colored knots and beads called *khipus*. The United States Federal Data Strategy will help support an open government through data standards and good practices.

One reason that the Ancient Inkas were able to manage enterprise data well was that they first standardized on objects such as person, place, or thing. Once these building blocks were in place data management was much easier.

The primary intent with the *Federal Data Strategy* is to "leverage data as a strategic asset". The strategy includes an action plan with twenty action steps. The author attempts to compare several Ancient Inkan techniques with Federal Data Strategy action steps as well as provide an *enterprise conceptual data model* (ECDM) visualization of the Federal Enterprise Architecture (FEAF) 2.0 Data Reference Model (DRM) standard for person, place, or thing.

AUTHOR BIOGRAPHY

Jim Watts is a Program Manager for the Transportation Security Administration, a component of the Department of Homeland Security, jim.watts@tsa.dhs.gov, linkedin.com/in/jim-watts-6b771b13 .