Avian Knowledge Network Data Sharing Levels

Michael Fitzgibbon
November 22, 2019

The Avian Knowledge Network works to leverage data, information, and partnerships to revolutionize avian conservation. In order to leverage data, we must describe and define clear, distinct sharing levels that express how data collectors and owners want to share their data with others. Our goal is to encourage data owners to share as many avian observations as widely as possible for conservation, with the understanding that each organization's needs are different.

As data is added to our systems, each record is flagged with a Data Sharing Level providing explicit control to data owners for how visible each observation is for querying, summarizing, visualizing, and analyzing without explicit permission from the data owner. Data Sharing Levels allow data owners to allow full access without restrictions, require a data sharing agreement for full access, allow data to be included in summary tools, only share metadata, or restrict sharing altogether. Data Sharing Levels are flexible, allowing data owners to, say, restrict access when observations are first recorded and broaden the availability over time as the sensitivity of the data diminishes.

Users manage sharing levels per observation visit, making all observations within a visit consistently shared as a unit, ensuring analytical integrity.

Working independently from Data Sharing Levels are the ability to discover sampling unit locations. For all applications that provide a geographic view of where observations are located (eg, AKN Observations Map), a flag on sampling units allows data owners to hide their locations from being discoverable.

The AKN Data Sharing Levels are broken into two stages: *In Review* and *Ready for Use*. During the *In Review* stage, data owners manage their project or organization data workflow for entering and proofing observations. Once the data are reviewed, the *Ready for Use* stage allows the data owner to choose how broadly they want to share their data.

As data moves between these stages and the levels within each stage, all AKN applications are responding accordingly, limiting access based on the person's role using the application as well as the specific Data Sharing Levels assigned. A data owner can change these Levels on data records at any time as they see fit. Note that once data has been shared openly and used outside of the AKN, changing those data to a more restricted level will only control future access.

The *In Review* stage levels are as follows:

In Review Level	Description
RAW	Data are freshly entered or uploaded, and are only available to Project members. RAW is the default level for all data added into the AKN.
CLEAN	Data that have undergone a first level of review, often by the person entering or loading the data, and are only available to Project members.
AVAILABLE	Data have been fully reviewed and deemed ready for the data owner to assign a Ready for Use stage Level (see below). AVAILABLE data are only visible to Project data owners.
RESTRICTED	Data have been fully reviewed, but the data owner does not want to share the observations at all. (Note that projects always share project names, descriptions, and contact information). No detailed metadata, data summaries, nor observation information is shared outside of the project.

Once data is AVAILABLE, the *Ready for Use* stage levels available are as follows:

Ready for Use Level	Description
METADATA ONLY	METADATA ONLY level shares the project description and contact, survey locations, survey methods, and the species surveyed. (Maps to AVAILABLE LEVEL 1 in AKN sharing levels v1)
SUMMARIZE ONLY	SUMMARIZE ONLY level inherits all METADATA ONLY rules, plus allow use of data as part of publically available visualizations and summaries that obfuscate observation details. (Maps to AVAILABLE LEVEL 2 in AKN sharing levels v1)
SHARE WITH PERMISSION	SHARE WITH PERMISSION level inherits all SUMMARIZE ONLY rules, plus allow use of full observation data and metadata detail to anyone that has a data sharing agreement with the data owner. (Maps to AVAILABLE LEVEL 3 or 4 in AKN sharing levels v1)
SHARE OPENLY	SHARE OPENLY level allows full observation data and metadata detail to anyone for conservation uses without the need for explicit permission from the data owner. (Maps to AVAILABLE LEVEL 5 in AKN sharing levels v1)

It's helpful to see exactly how the AKN applications will respond specifically to all Data Sharing Levels for all user roles:

Tool	Activity	Role	Data Location	Data Sharing Levels							
				In Review stage				**Ready for Use stage**			
				RAW	CLEAN	AVAILABLE	RESTRICTED	METADATA ONLY	SUMMARIZE ONLY	SHARE WITH PERMISSION	SHARE OPENLY
Bulk Uploader	add new datasets	Project Leader or Biologist	Projects	x							
Biologists	add new visit	Project Leader or Biologist	Projects	x							
Biologists	edit visit / observation data	Project Leader	Projects	x	x	x	x	x	x	x	x
Biologists	edit visit / observation data	Biologist	Projects	x	x						
Biologists	set data sharing levels	Project Leader	Projects	x	x	x	x	x	x	x	x
Project Leaders	add / edit project metadata (locations, methods, people)	Project Leader	Projects	x	x	x	x	x	x	x	x
Project Leaders	download datasets	Project Leader	Projects	x	x	x	x	x	x	x	x
Project Leaders	set data sharing levels	Project Leader	Projects	x	x	x	x	x	x	x	x
Data Sharing Level	set data sharing levels	Project Leader	Projects	x	x	x	x	x	x	x	x
Analysts	analyze observations	Project Leader	Warehouses		x	x	x	x	x	x	x
Phenology Tool (includes ECOS IPaC)	use	Public	Warehouses						x	x	x
Observations Map and node specific maps	use	Public	Warehouses						x	x	x
Data Downloader	find and download observations	Project Leader	Warehouses					x**	x**	x**	x
Data Downloader	find and download observations	Public	Warehouses								x

x** = available only for projects they are Project Leader on.

Roles listed above:

- Biologist - a registered user who can add and edit observation data in a project.
- Project Leader - a registered user who is the data owner for one or more projects who can add and edit project metadata and observation data and set **Data Sharing Levels** for their project(s).
- Public - an unregistered user.

Data Locations listed above:
- Projects - data managed and controlled by the assigned Project Leaders, supporting live editing, updating, and curation of metadata and observation data. Data in projects are organized to segregate user access and support organizational change control and allowing for a mixture of protocol types.
- Warehouses - project data automatically copied from Projects by the system nightly to support conservation application, including summaries, analysis, and visualization. Data in warehouses are organized by protocol types, supporting a mixture of projects.

One application that is not listed in this table is the Data Catalog, which supports an independent database that provides discoverable information about projects. A Project Leader can decide how much information to include or not in their Data Catalog records. Changing Data Sharing Level information on observation data has no impact on information available in the Data Catalog.