

BeeXML - The Promise of Standardized Data



Apimondia Working Group #15 (AWG15)
Standardization of data on bees and beekeeping

Robert Brodshneider, Ph.D

Joseph Cazier, Ph.D (Presenting)

Walter Haefeker

Peter Neumann, Ph.D

Marten Schoonman

Noa Simon, DVM

James T. Wilkes, Ph.D

Pim Van Gennip

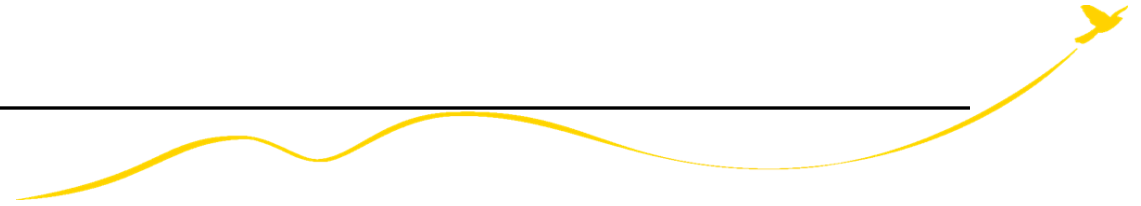


My Background

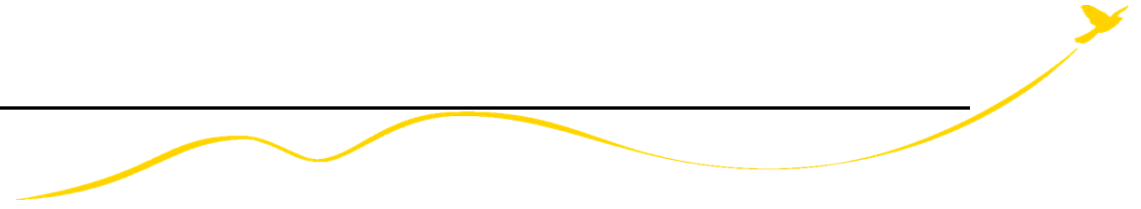
- Raised in a farm family with first degree in *Integrated Cropping Systems* focused on both the science (Agronomy) and precision agriculture (Technology) and how they relate in a system
 - Worked in both agriculture and research at various organizations
- Doctorate in *Information System* with a focus on *Data Science*
Currently Professor and Director of the Center for Analytics Research and Education - Focused on using big data to solve important problems
- Chief Analytics Officer for HiveTracks.com - Software for the tracking and management of bees
- Presenting on behalf of **Apimondia Working Group on Data Standardization**

Agenda

- Purpose
- Why we need data
 - Report and Policy Decisions
 - Smart Hive
 - Genius Hive
- Common Data Platform
 - BeeXML
 - Data Diversity
- Challenges/Opportunities
 - Privacy
 - Blocks for Bees
- Moving Forward



Why Bee Week



- We are all here because we know how important bees are and understand all too well the perils they face and how losing them would devastate our world
- This presentation is about a tool that can help us see clearly to better navigate those perils to help optimize the outcomes for bees and the crops that depend on them
- That tool, in the domain of data science, is analytics, which rests on big data

Data Science and Analytics

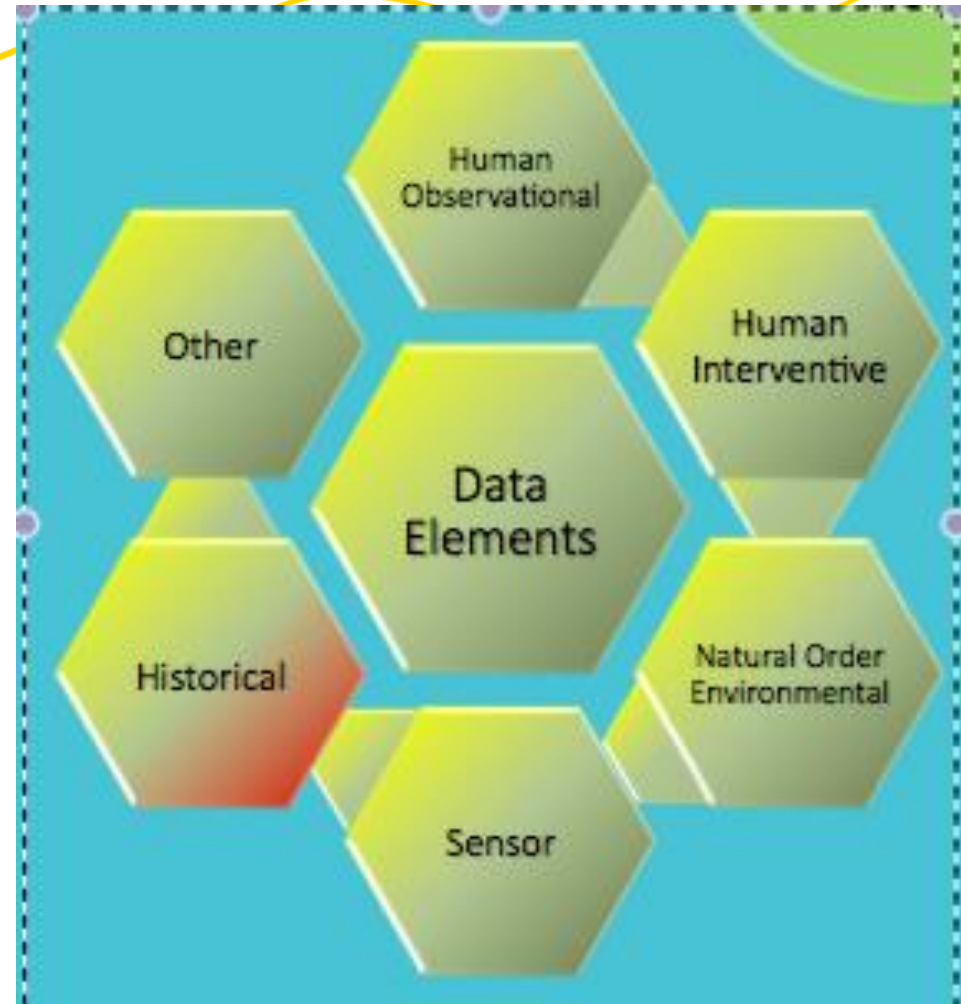
- Analytics is the process of making smarter decisions based on data
- We can see examples in related industries
 - U.S farmers that have implemented precision agriculture saw input costs decreased by 15%, crop yields increased by 13%, and water usage decreased by 50% (Bobkoff, 2015)
 - Farmers in Australia saved an average of \$14-\$30 per hectare using precision agriculture technologies. (Robertson, Carberry, & Brennan, 2007)
- Apiculture can see these types of benefits and more

Big Data

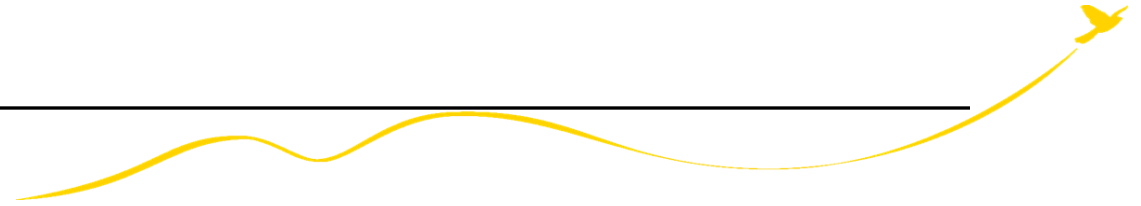
- Big data enables better more robust analytics
 - Quality, Quantity and Relevance all matter when it comes to data
 - Breadth and Consistency also matter
- What matters most is the **free exchange of data**, the ability to put relevant anonymized data together, and merge it with secondary data (weather, crop outcomes, etc) for deep analysis
- Can also allow for better reporting for policy makers and governmental decision making

Sources of Data

- Human Observational
 - Interventive (Treatment)
 - Natural World (Weather etc)
 - Sensors
 - Historical
 - Outcome Data
 - Others
-
- We need all of these sources of data.

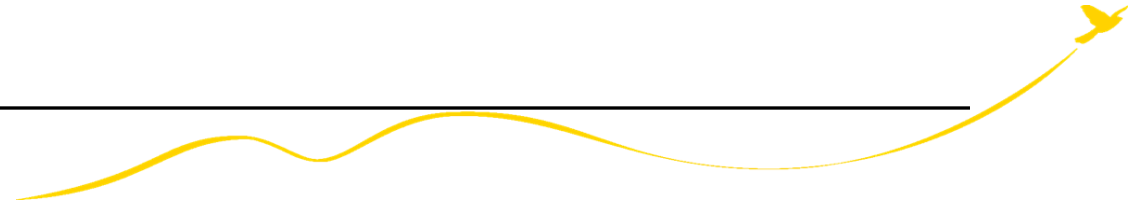


Smart Hive



- **Smart Hive** - A smart hive is one that can tell you about itself
 - Weight
 - Temperature/Humidity etc Sensors
 - Sound
 - Images
 - Locationally aware
 - Communicative
 - Temporal Dimension - remembers its past
 - Context sensitive - time of year, current and past management actions

Genius Hive

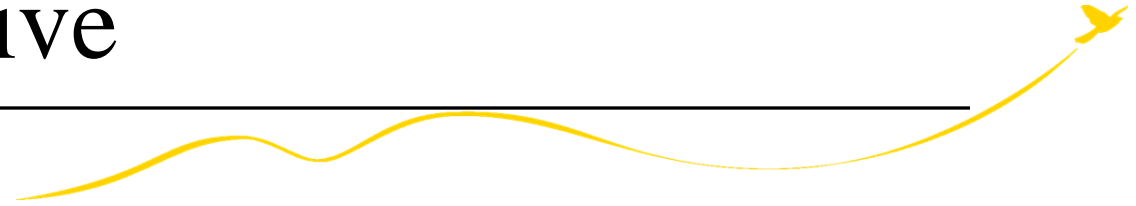


- **Genius Hive** - A genius hive is one that can tell you what it needs to be the best it can be
 - All of the information from a smart hive
 - + Years of standardized data that can be analyzed with *Machine Learning*
 - + Identification and verification of *Best Management Practices*
 - + **Software** to tie it all together

What Could a Genius Hive Do?

- Monitor the hive for signs of trouble and send alerts
- Monitor Regional and National Trends in Real Time
- Predict trouble *before* it happens
- Prescribe the best management *customized for a particular hive*
- Optimize health and performance - Going from *Good to Great*

The Path to a Genius Hive



- **Step One - Develop and Adopt a Data Standard**
- **Step Two - Identify and Validate Best Management Practices**
- **Step Three - Integrate Data Collection, Machine Learning and Best Management Practices into an *Intelligent Hive Management System***
- **Step Four - Continually use and improve the system with additional data and technology, sharing insights with the beekeeper community**

A Common Platform for Data Exchange

- In order to have Big Data to do important analytics on, we need a common data platform
 - Collect data from any type of software or other product
 - Have clear definitions and standards for that data
 - Allow for aggregating relevant portions of data for analysis or reporting
 - Have privacy protections
 - Be open for all to analyze and share knowledge from on an open learning platform

Layers of a Common Platform or Standard

- First layer is identifying the most important things to measure
 - Evolving process, start with more common data such as hive inspections, apiary suitability and management practices then expand
- Second layer is conceptually deciding how best to measure it
 - Once key area is identified as important agree on one or more standard measures for that - such as common problems
- Third layer is the technical architecture of how to store, transmit and analyze that data
 - The technical standard for storing and transmitting the data from machine to machine

BeeXML

- *Extensible Markup Language* (XML) has the potential to facilitate data standardization and proportional aggregation
- Similar to its twin, HTML (Hypertext Markup Language)
 - HTML is a standard for telling you what information should look like, mostly from **computer to a person**
 - XML is a standard for telling you what the information means, mostly from one computer to another computer - along with the meaning of the information

Potential of a Common Data Standard

- Just as HTML helped facilitate a computing revolution, XML has the potential to revolutionize beekeeping by allowing for the collection of large amounts of quality data
 - Scientific Reports
 - Citizen Science Efforts
 - Quality and outcome data
 - Live data from the front line of beekeepers
 - More data for better and faster analysis

Examples of XML



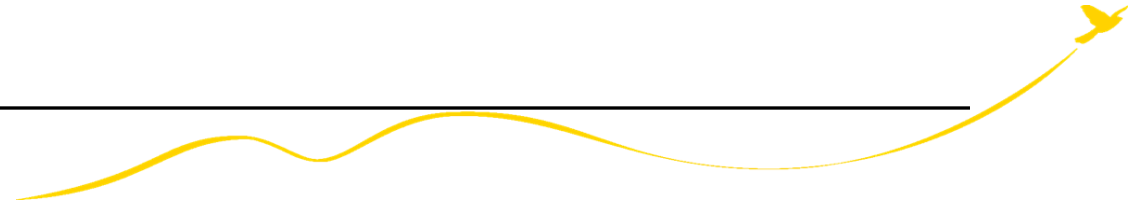
- Originally developed in the mid 1990s, in tandem with the rise of HTML, XML has become the default means of data exchange for many industries, including:
 - RSS, Atom, SOAP, SVG, and XHTML
 - Office Productivity tools (e.g. Open Office)
 - Industry data standards, e.g. HL7, OTA, NDC, FpML, MISMO etc. are based on XML
 - Is commonly used for data interchange in internet applications
- And many others. See sample code to the right.

```
<?xml version="1.0" encoding="utf-8"?>
<!--
- phpMyAdmin XML Dump
- version 4.6.6
- https://www.phpmyadmin.net
-
- Host: localhost
- Erstellungszeit: 20. Jun 2018 um 13:19
- Server-Version: 10.0.32-MariaDB
- PHP-Version: 5.6.34
-->

<pma_xml_export version="1.0" xmlns:pma="https://www.phpmyadmin.net/some_doc_url/">
<!--
- Structure schemas
-->
<pma:structure_schemas>
  <pma:database name="Sommersummen" collation="latin1_swedish_ci" charset="latin1">
    <pma:table name="objects">
      CREATE TABLE `objects` (
        `Objekt` varchar(255) COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
        `Beschreibung` varchar(255) COLLATE utf8_unicode_ci DEFAULT NULL,
        `Actions` varchar(255) COLLATE utf8_unicode_ci NOT NULL,
        PRIMARY KEY (`Objekt`),
        UNIQUE KEY `Objekt` (`Objekt`)
      ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_unicode_ci
        COMMENT='InnoDB free: 11264 kB';
    </pma:table>
  </pma:database>
</pma:structure_schemas>

<!--
- Datenbank: 'Sommersummen'
-->
<database name="Sommersummen">
  <!-- Tabelle objects -->
  <table name="objects">
    <column name="Objekt">Ableger</column>
    <column name="Beschreibung">Ableger</column>
    <column name="Actions"></column>
  </table>
```


Data Diversity

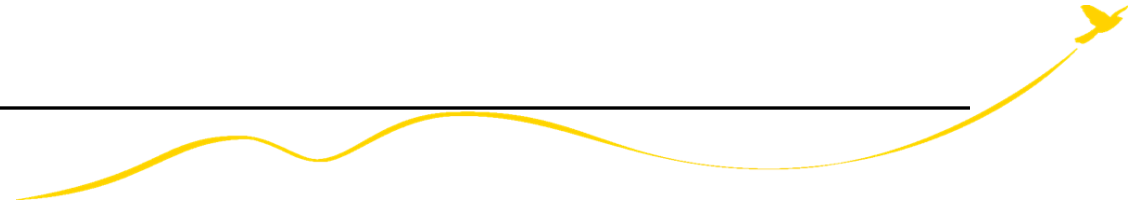


- The development of adoption of a common data standard like BeeXML would increase data diversity by
 - Allowing any vendor or organization of any data collection platform to merge their data into the system
 - Allow for free exchange of information along with its meaning
 - The inclusion of additional data over time and the standard grows

Challenges and Opportunities

- Opportunities
 - Analysis and reporting of data for better decision making
 - Development of the Genius Hive
 - All can learn from the anonymized data
 - Efficiency in data sharing
- Challenges
 - Commercial incentives are to own the data, likely leading to near monopoly control if one group owned all the data (e.g. Google), need to address
 - Agreement on that standard, and what is important to collect
 - Privacy Issues

Commercial Incentives



- A common data platform could be set up and maintained by the public sector, as a non-profit or NGO to collect and manage the data, making it available to researchers, governments and commercial organization under a common license agreement, with knowledge sharing back to the bee community
- By having a common standard like BeeXML, companies can focus on adding value from the data rather than monopolizing or owning it. Or by making and selling technologies to collect and act on the data

Privacy Issues

- Privacy issues could be addressed with Blockchain Technology that allows for anonymity and integrity of data
 - Data could be written to a blockchain distributed ledger with an encryption code that would verify its accuracy at the time of recording, leaving it immutable
 - The data could be aggregated and anonymized with private encryption keys, so you would know what was recorded, but not by who. Individuals could still access their data with the key
- If people know their data is private, they will share more and give better quality of data to the system, helping everyone

How the European Parliament Can Help

- Encourage Funding for technologically driven solutions such as:
 - Research and development to build, test, validate and host BeeXML
 - Research and development of blockchain technologies to address privacy and integrity issues with the data
- Legislation/Regulation
 - Establish a requirement that data from funded research be stored in this standard and stored in an accessible manner
 - Use this standard to collect and store bee data from other sources
 - Work with other governments and organization for the adoption and use of a world wide data standard similar to BeeXML

Questions / Discussion

Joseph@hivetracks.com

Or

cazierja@appstate.edu



HiveTracks
KNOW YOUR BEES