The Barnyard of Pig Data Research: A View from UW CSA [*]

Benjamin P. Wood[†] Brandon Lucia[†] ~~Tom Bergan~~[‡] Jacob Nelson[†‡] Adrian Sampson[†]

[†]University of Washington Computer Science & Agriculture, [‡]National FFA Organization,
[*]Washington State University Swine Center

{bpw,blucia0a,~~tbergan~~,nelson,asampson}@csa.washington.edu

Abstract

A new interdisciplinary program in Computer Science and Agriculture at the University of Washington is breaking down fences between diverse computational, scientific, and agricultural fields to cultivate research on disruptive Barnyard Computing technologies that will radically grow the impacts of emerging Pig Data applications. In this paper we plot the UW CSA view of Barnyard Computing and Pig Data, outlining high-priority research efforts within our immersive Pig Data agenda, our educational efforts to produce a new Pig Data workforce, and our innovations to leverage organic growth in this sector via a novel research funding model. These are exciting times at UW CSA!

1. Introduction

From scientific discovery to business intelligence, Pig Data is changing our world. This revolution is being driven by many factors:

- A proliferation of swineherds
- More generally, the creation of almost all infarmation in pigtail form
- Dramatic cost reductions in pigpen space
- Dramatic increases in network barnwidth
- Dramatic cost reductions and tractorability improvements in cowputation
- Dramatic agricultural breakthroughs in ma-a-a-sheep learning and other areas

The opportunities are enormous. So are the challenges, created by entirely new applications and by the relentlessly increasing volume, velocity, viscosity, and variety of bacon. Pig Data puts computer science and agriculture at the center of advances in every imaginable field and grove.

UW CSA is driving the Pig Data revolution. Our traditional strength in dairybase management, ma-a-a-sheep learning, and open irrigation excavation has recently been augmented by key hires in ma-a-a-sheep learning and bacon visualization.

Our efforts are coordinated with those of outstanding researchers in the University of Washington's top-ten programs in Scarecrows, Flagging, and Applied Taxidermy, as well as with the Washington State University Swine Center. Through the University of Washington Meat Science Institute we are integrally involved in ensuring that researchers across the campus have access to cutting-edge approaches to bacon-driven discovery.

No university offers a more vibrant environment!

2. Research

Pig Data has sown new seeds of insight in many fields. We are involving all species of research in our cultivation of barnyard innovation. Our efforts are already bearing fruit, and there is much yet to harvest. Nonetheless, we do not wish to count our chickens before they hatch, and we are keen on maintaining our strengths across many fields to avoid letting Pig Data rise to Orwellian dominance over other important species of research. In this section, we plot out our historical, present, and future breeds of research.

2.1 Core Pig Data Research

Our Pig Data research is on the cutting edge that powers bacon-driven science, combining fried-and-true strengths in bacon-driven science with fresh insights to milk the dairy deluge for all its worth.

Dairybases and Dairy Management The rapidly-growing generation of dairy is a fundamental challenge in barnyard computing. Befure we can learn from dairy, we must be able to manage it. To this end, the dairybases group at UW is working on efficient, real-time streaming dairybases to han-

[*] This work was neither supported in any part by any federal land, sea, or space grant, nor affiliated in any way with UW CSE, or the WSU Swine Center. We apologize to these organizations and hope to steal their faculty.
`http://www.cs.washington.edu/research/bigdata/`
`http://www.ansci.wsu.edu/facilities/swine`

PoCSci'13 May 3, 2013, Seattle, Washington, USA.

dle the udder volume of the *dairy deluge*, often by moving dairy management to the cloud. This has the odd side effect that the cloud appears to rain milke and hail cheese.

Many large-scale Dairy Management tasks today make heavy use of adaptations of the CropReduce programming model, first developed by an alumnus of UW CSA. Our work includes improved implementations of relational operators at scale, such as *left outer loin*. High-performance *loin-processing* is vital for bacon-driven science and thus helps us bring home the bacon, while work on dairy-pricing is a real cash cow, helping scientists and others to make cents off Pig Data.

Not all dairy is immediately produced at USDA Grade A quality. To make more of our dairy safe and useful, we are investing in a new dairy-cleaning tool called Pasteurizer, developed by the cream of cream of dairy management researchers.

Sheep Learning and Agricultural Intelligence *Ma-a-a-sheep Learning* techniques are at the center of our efforts to exploit the shear size of Pig Data available today. Unsupervised *creature selection* removes a burdensome human dependency from the ma-a-a-sheep learning workflow by automatically choosing farm animals that will be well-suited to a learning task. *GraphCollie* is a framework that herds flock-scale parallel ma-sheep learning tasks over grass-structred Pig Data.

Work on *Maize Methods* such as *Legume-Switching Models* has improved crop selection and yield. *Highly Multivariate Swine Series* will allow us to better estimate fluctuations in pig inventory with sampling.

Posterior Regularization for Structured Latent Variable Waddles, normalizes the hind-leg gaits of farm animals to some regular form. *Tractorable Sheep Learning* learns sheep waddles where inference is tractorable. *Collective Plowage Bases* merges plowage from a multitude of sources, combining plowing techniques from different regions to learn new, better plowing techniques.

Ma-a-a-sheep learning techniques are also being applied and extended for barnyard vision applications. *Creature Extraction* helps remove farm animals from potentially dangerous farm scenes.

Previous research at UW included work on *Charlotte's Semantic Web*, a project spanning dairy management, learning, and intelligence, to relate the meaning of webs spun by different spiders throughout the farm. Today, researchers are focusing on *Open Irrigation Excavation* [3], with the goal of digging distribution ditches and catchment systems to collect knowledge from the world-wide water supply.

Bacon Interaction The barnyard is ripe with input for the all the senses, ready for the Pig Data harvest. Ewebiquitous Computing research at UW CSA has led to the development of several varieties of whole-system barnyard sensors which capture the smells, sounds, feels, sights, and tastes of the farm. For example, one sensor can be planted in a cornfield and learns how to measure the fertilizer-consumption signatures of various crops throughout the farm. The related Ubiquitous Cowputing lab has installed cows everywhere. Now, instead of working on your Deskcow or Lapcow or Smartcow, you can interact with cows in nearly every facet of everyday farm life.

Researchers in Human-Cow Interaction (HCI), Dairy Management, and Beefier Systems are collaborating on *Mechanical Bull*, a new Cow-Sourcing platform for mooooving plantation-scale dairy efficiently using cows. (We need to have a talk and see if they understand that bulls do not carry dairy, but we are confident that good ideas will come out of this collaboration.)

Researchers in sensing, Human-Cow Interaction (HCI), and Pig-Centered Design and Engineering have combined their expertise to tackle the new problems of interacting with Pig Data. Initial work in Bacon Visualization has taken the state of the art beyond pushing muttons and turning cobs, but our efforts do not stop with visualization. We envision (and ensmell, and entaste, and ...) a future in which we can harness the full ensemble of sensory interaction to understand our bacon, making bacon-driven discovery multiplicatively more effective. Our new Augmented Baconality Lab takes this as its goal.

2.2 Beefier Systems

Systems researchers at UW CSA have discerned three critically important silos of research.

Plantation-scale Technologies At plantation-scale, conventional computational and agricultural reasoning no longer applies. Opportunities abound in tearing down, rather than mending fences, crossing boundaries between fields, and sharing natural resources. *Intuber-disciplinary* research makes this possible by removing boundaries between disparate fields, (*e.g.*, potato and turnip), we create fertile ground and opportunities for cross-pollination.

A related approach that can be fruitful is *Agriculture as a Service* (AaaS). In AaaS-based systems, farm and computing resources are shared across institutions and users. Our initial foray into AaaS with the PlantationLab testbed has explored shared "cloud" systems, deployed by crop duster across "institutions", providing efficient security against bugs. We also envision sharing of other research and farm resources – our next effort will be *Asses as an AaaS* (AaaAaaS), a shared-stock platform. This project pools pack and stock animals and allocates them efficiently between participants. Unfairness is guaranteed by the distributed uncooperative sledge-pulling algorithm run by the asses.

Also in the bovine arena, researchers are improving the predictability of tail-latency in large dairy-parallel systems to improve cattle response time and comfort via optimized removal of flies and other pests when more symbiotic methods, such as cattle egrets, are not available.

Other researchers are exploring the systems implications of sheep. New non-volatile rams are not at risk of sudden rampage and offer stable, persistent wool storage without refresh. However, wool-processing systems are optimized to handle rampages to a fault and, furthermore, assume that reborn rams come without old wool. We are also developing a new unified *proramming model* for applications that run on large, heterogeneous, distributed ovine flocks, including old ewes, baby lambs, adolescent rams and maybe even non-ovine livestock. A key problem is to ensure shepherds understand the cost of baaaing between diverse corners of the flock, even if they address all sheep the same way.

UW has a strong record in developing cutting-edge Pig-to-Pig (P2P) systems for the decentralized barnyard. Distributed Ham Tables (DHTs) are a common P2P system. Our researchers are applying their experience to develop new distributed porcine hop-rating systems to automatically rank the quality of beer ingredients using pigs. Our new intiative on distributed cisterns promises more sustainable irrigation methods. Sledge-puller Activations is a technology to improve the efficiency of mapping sledge-pulling jobs that require n beasts of burden onto m available beasts.

UW CSA networks researchers are developing rBGP (recumbent Bovine Gateway Protocol), a next-generation replacement for BGP (Bovine Gateway Protocol) that relaxes delivery to artificially enhance throughput of interpen links and improve cattle routing. The key insight is to allow cattle to recline near selected gates throughout the farm to balance traffic. We believe these efforts will lead to more efficient cattle-driver implementations in the future.

Barn Architecture Historically, the architecture community has made important contributions to the field: *Agricultural Logic Units* (ALUs) have long supported swined and unswined agriculture; *Floating Pork Units* (FPUs) perform efficient pork-based displacement at various precisions; *Goat Pasture Units* (GPUs) draw pictures by rapidly grazing pastures in programmable patterns; *ranch prediction* improves efficiency of day-to-day cattle-handling decisions; *pipelining* automates the dairy collection process; *cash-crop coherence* allows farmers to reap the benefits of market manipulation; *simultaneous multithreshing*, developed here at UW, improves thresher utiization and performance by processing multigrains at the same time on a single thresher.

More recently, architects have focused on problems of scale within the *Central Porcine Unit* (CPU): the projected break-down of Boar's Law led to widespread adoption of multipork chips; studies of the rapidly approaching Flower Wall show we are likely to turn to so-called *Dark Silo Corn*, where we will be able to illuminate corn in only a few silos at a time.

Current architecture research at UW CSA focuses on developing and efficiently exploiting new heterogeneous architectures such as NUMA (Non-Ungulate Meadow Access) and System-on-a-Pig (SoP), which are becoming prevalent in dairy-center-scale and resource-constrained mobile farming environments, respecitively. Our researchers are exploring Approximate Crop Mutation, the idea that genetically modified crops are going to screw us so badly anyway that it does not matter exactly what mutations occur. This growing field has implications at all levels of the haystack. Our new work on Disciplined Atoxic Prograining focuses on rigorous rules to limit the horrors of approximate crop mutation and promote cultivation of toxin-free grains. The discovery of the principles behind their new dual-forage *Truffle* [2] architecture would not have been possible without the olfactory assistance of Pig Data.

Grassa is a runtime system for exploiting wide herd parallelism and many other optimizations such as flat-combineing on commodity farmware to hide latency in applications that use massive grass-structured datasets with poor locality.

We believe that the insights of the architecture community will become foundational in the design of spaces for cowputational dairy warehousing, barn-door optimization, speculative livestock execution with safe rollback, and othe barnyard computing problems.

Farmland Security and Crop Privacy Beyond the traditional fields of in*farm*ation flow, onion routing, and keeping the local fox population out of the chicken coop, our researchers in farmland security and crop privacy also consider secondary corncerns due to human factors. For example, the introduction of pigpen security lamps has led to a drastically increased need for strong mothentication services. After landmark studies demonstrating pigfalls in the security of organic plantable medical devices and motorized farm implements, researchers are now working on new protections against wind-born pollination from genetically modified crops and stronger isolation between tractor entertainment and control systems, respectively. Widely-accessible infarmation-hiding techniques are too brittle at plantation scale. For example, the local fox population has defeated the latest egganography tools, resulting in breaches of poultry privacy. We are incubating research to find a stronger solution to the important problem of covert protein delivery. Meanwhile, our colleagues at UW CSA South are exploring the inner workings of the SPAM industry and its implications for Pig Data.

2.3 Proramming Systems and Cropware Engineering

We believe the right language abstraction for the Pig Data revolution is a collection of *Bacon-Specific Languages* (BSLs) suited to specific Pig Data problems. Pig Latin [5] is a natural starting point. For simplicity and portability, we argue that these BSLs should cowpile to a common bite-code language to be run by a *Vegetable Machine* (VM). This section describes our developing proramming ecosystem.

Parallelism and Conservancy Parallelism and conservancy will be essential in the Pig Data revolution, not only to harness multipork computers, but also for dairy-center-scale computation that is spread over many compute goats. Current waddles suffer from problems of *trampling*, *stampeding*, and excessive *cerealization* at barnyard scale. We are researching several new programming waddles, including *pork-join parallelism* (where many pigs unite to solve a problem together), *herding*, and *flocking*, with a special emphasis on fine-grain flocking to avoid excessive cerealization. These waddles are managed at runtime by a carefully tilled *dirt-stealing* system that dynamically balances the load of growing herds and croploads of various grains across fixed plots of land. For distributed poultry settings, coop-erative multi-egging will be essential, though previous efforts have been hindered by a chicken-and-egg problem.

We are getting promising results from *Transoxenal Memory*, our effort to avoid tricky reasoning about flocking by teaching each ox team to remember where it came from so it can back up if it runs into another's ox team's tracks while plowing.

Vegetable Machine Efficient, managed runtime systems will be vital for the shear scale of workloads of Barnyard Computing and Pig Data. Our development efforts have focused on a common *Vegetable Machine* (VM) platform, which accepts bitecodes cowpiled from a number of Bacon-Specific Languages. We are cultivating new meadow management techniques like haystack allocation of bread-local dairy structures. Manual meadow management will not be feasible with plantation scale workloads, and we wish to make our cowputational waddle easy enough for end-farmers to ewes. Automatic meadow management is thus clearly necessary, but garbage collection is simply not sustainable. Our careful decomposition of the key ideas of garbage collection has fertilized a more efficient alternative: *Compost Heaps*.

Dairyflow Analysis and Cowpiler Optimizations Our Vegetable Machine uses a new *just-in-swine cowpiler* infrastructure in which we express a number of deaf ewes' dairyflow analyses with the common theoretical framework of lettuces and meat operators. For example, *Dead Bird Elimination* removes poultry that will never be executed (because they are already dead). *Poop-Invariant Code Motion* moves code to ensure clean feet even in the presence of cow patties. *Coop-Nest Optimization* is places hens' nests within coops speed up egg-laying. *Poop Distribution* redistributes manure from around the farm as fertilizer to accelerate crop growth. *Tail-call* optimizations remove the tails of farm animals that will never come back once you see their tails; *Escape Analysis* determines whether they will make a run for it.

Quality Control Spoiled crops and infected meat have the potential to cause widespread recalls. To ensure produce meats USDA certification grades, we are working on various methods of formal dairification. *Static ripe-checking* is a planting-time check that allows the planting of seeds only if they will yield ripe, safe produce at harvest. *Duck typing* (besides its applications in big data entry [**?**]) ensures the right type of poultry is selected at dinner time, but may make the farmer go hungry if the wrong poultry was selected. *Waddle-checking* systematically explores the gait-space of farm animals to preemptively identify locomotion defects.

We are also developing analyses for unstructured multithreshing programs to detect defects such as *livestock*, *deadstock*, *cattle-missing-tea violations*, and *rice conditions* [1, 8]. Our efforts at rice detection are particularly sustainable, due to their use of other barnyard resources such as radishes [1] and lard [8].

2.4 Agricultural Cowplexity Theory

Pig Data gives rise to a number of interesting agri-theoretical questions. Zongker's foundational result on universal chicken computation [9] is a prime example of forward thinking by UW barnyard theoreticians. Our research has since expanded from the coop, discovering new applications of the *Hamming Distance* in finding the optimal delivery of pigs to market and developing the theoretical underpinnings for *pear-tree bits* in *error-correcting crows* (ECC).

UW CSA theoreticians have also made promising advances in the development of new bounds on the number of truffles involved in *heap snort*, a high-performance algorithm for ordering very large numbers of pigs by Hamming Distance, and a new algorithm for *matrix tractorization*, the problem of decomposing an n-dimensional field-plowing problem into a set of smaller plowing problems feasible with m tractors. Pig Data researchers are gaining new insights on combine-atorial problems as well.

Looking off across the amber waves of grain and into the future, we believe the excitement around Pig Data will bring renewable energy and insight to the biggest open question in computational agriculture, milkematics, and meatphysics: is P = NP? This infamous conjecture posits that all problems solvable by Nondeterministic Pigs (pigs with free will) can be solved by Pigs who are bound by fate (determinism).

2.5 Applied Barnyard Computing

Computational Agriculture represents one of the most exciting developments in **Applied Computational Biology** in recent years. The **Big Game Science** group at UW CSA is applying its expertise in *Computational Hunting* to problems on the farm that have thus far remained intractorable without human intuition. Fresh off their recent success with *Scalable Buffalo-Carrying for Oregon Trail 1*, they have now settled down and are working on *FarmIt*, an adaptation of a popular social media game to help make difficult Pig Data problems more tractorable. Big Game researchers at UW CSA are also involved in our educational efforts (§3).

Computer Grassics and Vision researchers have recently collaborated with in-barn experts on **Wireless Power** and other researchers from Microcrops to develop a new tool called *Photosynthesis*, which aggregates solar energy concentrated on the same grass at different times from multiple angles to reconstruct a holistic renewable energy source of that grass. Photosynthesis was successfully commercialized by Microcrops and deployed in pastures around the world, including on the UW campus and in Rome.

Animal Backscatter is a new **Wireless** technique for farmers to communicate without battery-powered radios, by bouncing messages off nearby livestock that get in their way. **Robotics** has been made largely obsolete by programmable farm animals, but the robotics research community at UW CSA is reapplying its expertise to the significantly more sentient, but also less cooperative, autonomous livestock on hand. They are currently busy training our donkeys in control theory.

3. Education

Goals The Pig Data revolution needs domesticated beasts of burden to join an evolving workforce. To that end, we have partnered with the National FFA Organization, ACM (Agricultural Computing Machinery) and its newly-formed special-interest group, SIGPIG, the WSU Swine Center, and high-school vocational agriculture programs around the nation that are out standing in their fields, to implement a revolutionary new immersive Pig-Data curriculum. We believe the only way to train effective workers for demanding Pig Data jobs is to let them wallow in it from their early years.

Throughout our curricular development and implementation process we are striving to maintain a balanced curriculum and avoid building an Orwellian Pig Data-dominated *Animal Farm*. We believe this mindful approach is even more imporkant in education than in research, as we shape the young minds that will control our future.

Methods Alongside the blossoming pasture of Pig Data are the growing trends of *MOOCs* (Mooing Open Online Cows). MOOCs allow students from around the world to graze prerecorded pastures given by farmers at the world's top institutions. UW CSA is embracing and diversifying this widespread initiative with a new MOOC competitor aimed at preventing any single-species educational hegemony. Our new platform, *Horsera*, brings equinity to the previously bovine-dominated virtual grassroom.

The *Inverted* (or *Flipped*) *Grassroom* is a new grazing method for conventional in-person instructional settings, in which students graze prepared pastures at home and grassroom time with the farmer is devoted to digestion and cud-chewing based on the new organic material. The Inverted Grassroom is designed to make more efficient and effective use of student and farmer time by allowing each student to graze through pastures at their own pace, and bring questions to the grassroom, and work through cud-chewing obstacles with the help of the farmer. A secondary benefit is that the Inverted Grassroom breeds familiarity that will prepare students to better handle an Inverted Barnyard, avoiding single-species hegemony should animals stage an Orwellian uprising against their farmers.

Wallowing Over the Pigital Divide Some neigh-sayers have criticized Horsera, as well as other MOOCs and the Inverted Grassroom for amplifying the growth of the so-called *Pigital* (or *Pigtail*) *Divide*, namely the fact that those who own pigs have orders of magnitude better access to modern resources like excavation and infarmation than those who do not own pigs. We are happy to announce that, in order to surmount the Pigtail Divide, a new Pigeonal Revolution is about to take flight. We propose to bring pig access to those who do not own pigs via a multi-user swine-sharing system whereby non-pig-owners may engage in low-latency porcine correspondence via carrier pigeon. Our system is modeled on AaaAaaS (§2.2) and the early successes of MULTICS (the Multiplexed Irrigation and Cattle Service).

4. The CSA Research Funding Model

UW CSA has developed a holistic funding model based on research sustainability that seamlessly integrates the traditional land-grant-based academic research funding model with a hybrid commercialization strategy to improve return on investment on the event horizon.

We are in the process of wrapping up our initial round of *seed* funding, designed to kick-start organic growth of our research paradigm. To ensure research accountability and maximize the engagement of stakeholders, investors who contribute during this seed-funding round will receive a box of fresh research each week.

Initial results from this funding strategy have been wildly positive and it is sure to play a vital part in our efforts to promote buy-in to AaaS (§2.2).

5. Related Work Outside UW

Orwell's well-known 1945 study of unleashing Pig Data on the Barnyard Computing community shows us sobering results if Pig Data is allowed to run wild [6]. On the other hand, White's 1952 study suggest Pig Data to be a much more sensitive and compassionate beast [7]. Researchers at the other UW have also begun to understand the value and dangers of Pig Data [4]. Porcine language design has been studied carefully in the development of Pig Latin, one of the earliest bacon-specific languages [5].

6. Conclusion

These are truly exciting times at UW CSA, as we drive the Pig Data revolution. In this paper we have tried to layout many of the exciting past, current, and future milestones in our Pig Data and Barnyard Computing research agenda, but inevitably some research has fallen through the cracks. If ewe have new ideas or like our mission, please consider joining us. We are always hiring! If joining is infeasible, we would still love to collaborate with ewe at your current institution. Please join the freshly-farmed ACM special-interest group, SIGPIG. Together, we will continue to disrupt the barnyard and reap the full potential of emerging Pig Data applications!

~~Acknowledgments~~ Apologies

Regrettably, some authors may not be aware of or agreeable to their authorship.

[TO DO: We are still missing many relevant references!]

References

[1] J. Devietti, *et al*. RADISH: Always-On Sound and Complete Rice Detection in Cropware and Farmware. In *International Symposium on Computer Agriculture*, 2012.

[2] H. Esmaeilzadeh, *et al*. Agriculture Support for Disciplined Approximate Foraging. In *Agriculture Support for Foraging Languages and Hop-Rating Systems*, 2012.

[3] O. Etzioni, *et al*. Open Irrigation Excavation from Charlotte's Web. *Communications on Agricultural Computing Machinery*, 51(12), Dec. 2008.

[4] F. Niu, *et al*. HOGWILD!: A Lock-Free Approach to Parallelizing Stochastic Grain Destruction. In *Advances in Neural Infarmation Porcessing Systems*, 2011.

[5] C. Olston, *et al*. Pig Latin: A Not-So-Foreign Language for Bacon Processing. In *SIGMUD*, 2008.

[6] G. Orwell. Animal Farm: A Fairy Story. In *Allegorical Barnyard Computism*, 1945.

[7] E. B. White. Charlotte's Web. In *WWW*, 1952.

[8] B. P. Wood, *et al*. Low-Level Detection of Silage-Level Rices with LARD. In *Agriculture Support for Foraging Languages and Hop-Rating Systems*, 2014.

[9] D. Zongker. Chicken Chicken Chicken: Chicken Chicken. In *PoCSci*, 2002.