QDR
QUALITATIVE DATA REPOSITORY

Orchestrating Cloud Infrastructures to Manage Sensitive Data

IDCC 2017, February 20-24, Edinburgh

10.6084/m9.figshare.4670533

Nic Weber, University of Washington

Sebastian Karcher, Qualitative Data Repository

Alex Ivanov, Qualitative Data Repository

Sebastian Ostrowski, Qualitative Data Repository

QDR Overview

- At Syracuse University (United States)
- Online since 2014, NSF funded
- Curated, qualitative & mixed methods data
- Small holdings (~20 datasets), growing steadily

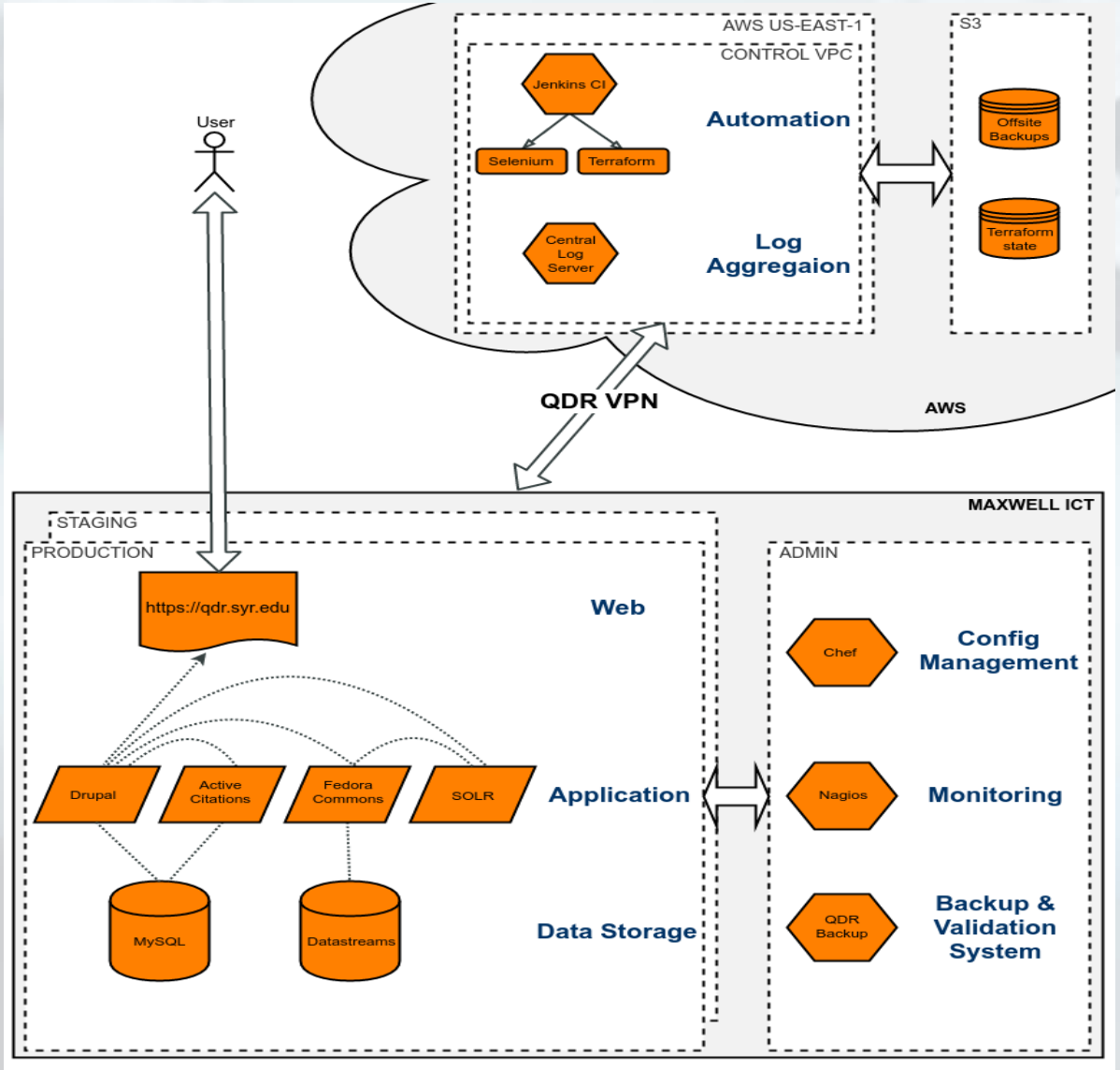
QDR

Why Cloud Infrastructure

- Elastic Capacity / Load Balancing
- Single platform (some services already on AWS)
- Comfort of built-in tools / easy to set-up automation
- Redundant storage
- Economical
- **But:** Must allow for storage of sensitive data

QDR

QDR's Current Infrastructure



AWS Cloud

ICT Syracuse

QDR

Sensitive Data, Cloud Storage, and the Importance of Context

- US Legal Context: Patchwork of laws
  - Electronic Communications Privacy Act (ECPA) (which is composed of the Wiretap Act, the Stored Communications Act, and the Pen Register Act); Computer Fraud and Abuse Act (CFAA); Health Insurance Portability and Accountability (HIPAA) which includes electronic Personal Health Information (ePhi), and the Health Information Technology for Economic and Clinical Health Act (HITECH Act); Children's Online Privacy Protection Act (COPPA), and Family Educational Rights and Privacy Act (FERPA)
- But overall permissive privacy laws
- QDR Context: Small team and holdings

QDR

CAIQ Self-Assessment

- Consensus Assessment Initiative (2014)
- Catalog of 292 Yes/No Questions
- 16 Areas, e.g. Application & Interface Security, Datacenter Security, Encryption & Key Management
- AWS recently published own assessment: https://d0.awsstatic.com/whitepapers/compliance/AWS_Risk_and_Compliance_Whitepaper.pdf

QDR

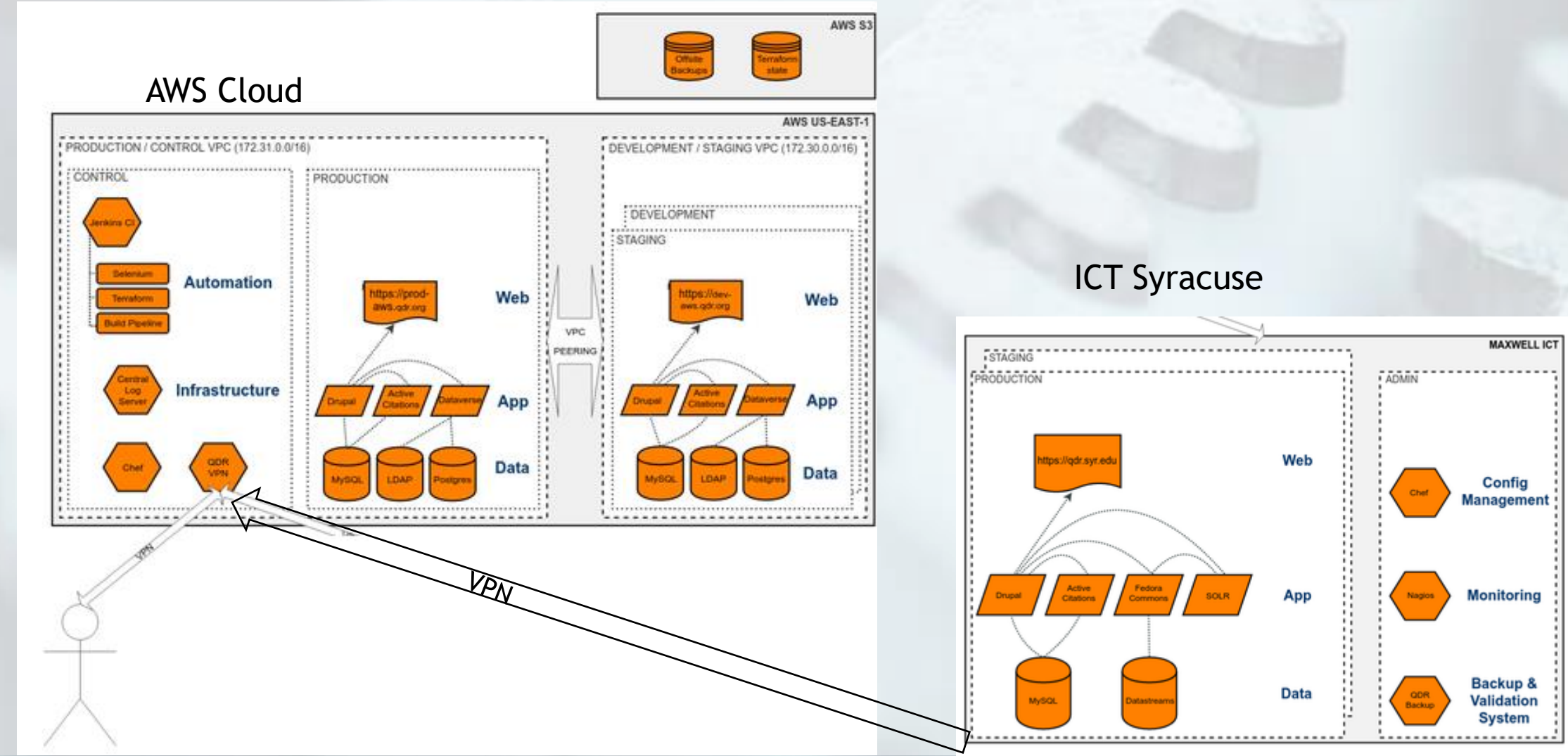
Using CAIQ Self-Assessment for QDR

- Transforming VPN set-up to use VPC
- Single Sign-on to both AWS and Syracuse ICT
- Shared online at: https://doi.org/10.6084/m9.figshare.4670506
- Example:

Network architecture diagrams shall clearly identify high-risk environments and data flows that may have legal compliance impacts. Technical measures shall be implemented and shall apply defense-in-depth techniques (e.g., deep packet analysis, traffic throttling, and black-holing) for detection and timely response to network-based attacks associated with anomalous ingress or egress traffic patterns (e.g., MAC spoofing and ARP poisoning	Do your network architecture diagrams clearly identify high-risk environments and data flows that may have legal compliance impacts?	x			This is something we should discuss in future… how to represent this compliance in our system's diagrams	S3.4
	Do you implement technical measures and apply defense-in-depth	x				

QDR

QDR's Planned Infrastructure



Benefits of Self Assessment

- Assessment of AWS Suitability (low sunk costs)
- Early recognition of config issues rather than during testing
- Recognition of additional documentation/policy requirements

QDR

NIST Cloud Usability Framework

- Ongoing works
- High-level framework for any cloud service
- 5 Categories: capable, personal, reliable, valuable, secure
- Repository-user as end user: user stories
- Developer/organization as end user

QDR