UN Global **Trusted** Platform
Data, Applications and Services

Proof of Concepts
Approach

• Minimal Viable Team
  • 2 x Technical Resources
• Wardley Maps
• Principles
• Configure / Show / Iterate
• Cloud Computing
  • Consumption Based Pricing
  • Commodity Infrastructure
• Agile Delivery
  • Working with users
  • Prototyping
  • Iterative Development
Research

- 160+
  - Platforms
  - Products
  - Services
- What did we find
  - Data
  - Partners
  - Domain Knowledge
  - Methods
- Trust

Data
- NSIs
- Open
- Proprietary
- Personal
- Geospatial
- Social Media

Partners
- NSIs
- Companies
- Academia
- Individuals

Domain Knowledge
- NSIs
- Industry Specific
- Agriculture
- Finance

Methods
- NSIs
- Proprietary
- Industry Specific
- Academia
Data Collaborative

- Method
- Data
- Partners
Encourage innovation via direct action, collaboration and self-service

Use ecosystem to identify future patterns
Proof of Concept
Wardley Maps

https://medium.com/wardleymaps/on-being-lost-2ef5f05eb1ec

https://www.youtube.com/watch?v=LZrINICJxUY
The process of evolution repeats...
Gaps / Next Steps

• Standards
  • Trusted Data
  • Trusted Methods
  • Trusted Partners
  • Trusted Learning

• Data Policy Framework

• Common Technology Infrastructure
Gaps / Next Steps

• Build Marketplace/Accelerator
  • Use to test business model/case
  • Collecting intelligence on data collaboratives
  • NSIs
Procurement

This page is intentionally left blank

Amazon AWS Costs

- July $30
- August $480
- September $683
- October $1,664

Total $2,857
Proof of Concept
Demonstrations
Proof of Concept
Apache NiFi
Apache NiFi

- CSV/Excel/File Processing
- Data Validation/Transformation
- Social Media – Twitter
- Satellite Data
- Integration with AWS Athena
- Integration with Solr/Banana
- Web Scrapping
- Public APIs

- Apache NiFi
Apache NiFi

• Open Source
• Niagara Files (NyFy)
• Highly Scalable
• Cluster Support
• Cloud Agnostic
Proof of Concept
Material Properties of Data
Material Properties of Data

- Data Ingest Workflow
- Antivirus Check
- Preview Data
- Classify Data
- Material Properties
- Ingest Data
Material Properties of Data

- Data Ingest Workflow
- Apache NiFi
- R
- RStudio
- Shiny
Proof of Concept
Satellite / Vector Data
Satellite / Vector

- UK Buildings Shapefiles
  - All UK Buildings
- Global Database of Events, Language, and Tone (GDELT)
  - Imported from AWS S3
  - Imported 6M Events
Satellite / Vector

• Open Source
  • GeoMesa
  • GeoServer
  • GeoWebCache
  • Spark
  • Accumulo
  • Hadoop
• AWS EMR
Proof of Concept
Satellite / Raster Data
Satellite / Raster

- Find, combine and analyse earth observation data
- Browse existing satellite datasets
- Stitch together imagery
- Build analyse automated pipelines
- Edit, iterate quickly (Real-time)
Satellite / Raster

• Process large/small raster datasets with low latency using multiple clusters and multiple threads
  • GeoTrellis
  • GeoMesa
  • Spark
  • Accumulo
  • Hadoop
  • Scala
Proof of Concept
Integrated Development Environment (IDE)
Integrated Development Environment

- Methods
  - Create / Share Code
  - Share Data
- Manage Dependencies
- Collaboration
- Windows / Linux / Macs
- Local / Cloud
- Assured Packages
Integrated Development Environment

- Anaconda 2/3
- JupyterLab
- Shiny
- Supports R, Python
Proof of Concept
Methods Library
Methods Library

• Easy to use
• Easy to find
• Easy to compose
  • Building blocks
• Add to workflow
Methods as a Service

- Access via APIs
  - Methods as a Service
- Documentation
- ISTAT R Package
  - R Evolved Generalized Software for Sampling Estimates and Errors in Surveys

API

- Simple utility functions
- Higher-level methods
- ML and deep learning models
- Advanced, use-case specific solutions

Increase in abstraction
Example Services
Example Services/Products/Code

• Kylo
• GeoTrellis
• Australian Open Data Cube
• GeoNotebook
• Rstudio Server
• Cartoview
• Vinsight

• Multi-Party Authentication
• GeoNode
• OpenMined
• DataShield
• rSDMX
• FarmShot
• Plantix
Useful Links

• GitHub Repository
Additional Slides
Proof of Concept
Calculations on Encrypted Data
Calculations on Encrypted Data

- DARPA Project
- Jana Database
  - Open Source
- Encrypted At Source
- Multi-Party Computation
- Differential Privacy
- Homomorphic Encryption
- High-Tech
Jana Capabilities

- **Functionality**
  - Generous subset of SQL
  - RDBMS Atomicity, Consistency, Isolation, Durability (ACID) properties

- **Privacy**
  - Data in transit - Public key crypto + proxy re-encryption
  - Data at rest - Deterministic, random, searchable
  - Computation - in RDBMS using Deterministic Encryption (DET) & searchable, in SPDZ MPC
  - Results - Differential privacy applied while in Multi Party Computation (MPC)

- **Performance**
  - 10Ks of records moving to 100Ks, queries in seconds

- **Deployment**
  - Virtualized appliance with RESTful API
Types of Queries

- SPJ, UNION, INTERSECT, EXCEPT
- Integer, String, Boolean, Enum, Fixed-Point, Date
- Nested queries

*Video Link*