GODAN ACTION
Global Open Data
for Agriculture & Nutrition

GODAN Action WP1: data standards: survey, gap analysis and recommendations

Focus on weather data standards

Valeria Pesce
Global Forum on Agricultural Research and Innovation (GFAR)

Leigh Dodds, Jeni Tennison, Pauline L'Hénaff
Open Data Institute (ODI)

Panagiotis Zervas (AgroKnow)

Global Open Data for Agriculture and Nutrition initiative (GODAN)

GODAN
Global Open Data
for Agriculture & Nutrition

The GODAN initiative was one of the outputs of the **G8 International Conference on Open Data for Agriculture** and was announced at the Open Government Partnership Conference in October **2013**.

The initiative focuses on **building high-level support among governments, policymakers, international organizations and business**.

Currently around **500 partners** worldwide from national governments, non-governmental, international and private sector organizations.

www.godan.info

GODAN Action

GODAN ACTION
Global Open Data
for Agriculture & Nutrition

Three-and-a-half-year programme launched by the UK's Department for International Development.

Under the DFID GODAN funding stream

UKaid
from the British people

GODAN Action brings together agriculture and nutrition specialists and open data experts and will support GODAN in its mission by building **people's capacity to engage with open data**.

http://www.godan.info/godan-action

GODAN Action: Focal areas

1) **Standards** - Enhancing data standards and promoting best practice in agriculture and nutrition to improve interoperability.
    1) Map of agri-food data standards
    2) Gap analysis on use and usability of data standards
    3) Recommendations to address gaps
    4) → Specifications → services, pilot implementations

2) **Research** - Identifying and improving tools and methods for evaluation of the impact of open data usage in initiatives and investments in agriculture and nutrition.
3) **Capacities** - Building the capacity and diversity of open data users, leading to more effective use of data in tackling key agriculture and nutrition challenges.

Purpose of the global map of data standards

- The main purpose of a global map of data standards in a specific field *is to **promote the discovery and reuse of vocabularies** and their properties, classes and controlled values. The reuse of existing vocabularies **promotes greater interoperability between vocabularies and datasets**.*
  paraphrasing what the Dublin Core Metadata Initiative says about their DCMI Registry (http://dcmi.kc.tsukuba.ac.jp/dcregistry/)

- Help **identify overlaps, duplication, gaps** and limits to adoption,
  >> encourages not to duplicate efforts and to collaborate to both develop and use common standards

Approach for building map of data standards

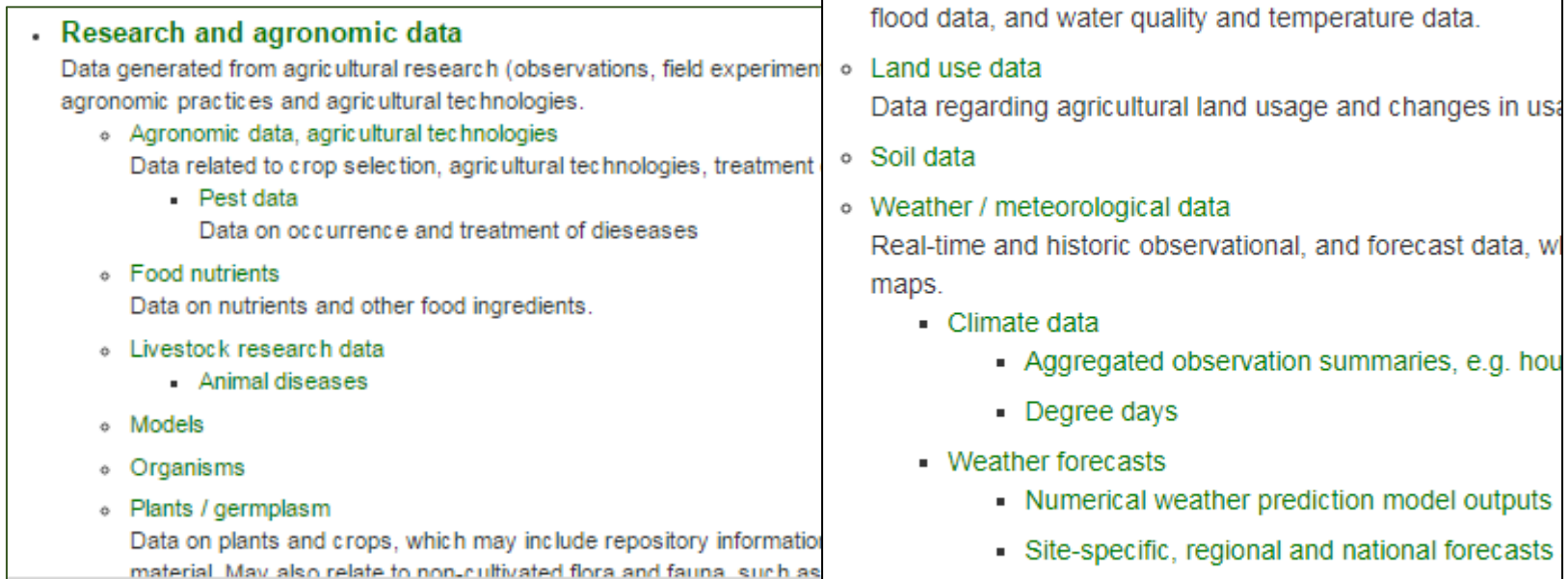
- **Not duplicating**, building on what exists (e.g. sync with AgroPortal)
- **Collaborative effort**
  - **Call to action to partners and experts**
  - **Anybody can add or claim a standard**
  - **All contributors acknowledged**
- **Broad coverage**
  - Agronomy
  - Natural resources
  - Fisheries
  - Value chains
  - Ontology
  - Thesaurus
  - Taxonomy
  - Code list
  - ISO specification
  - Messaging standard
- **Open data angle**
  - Format
  - APIs
  - Mappings
  - License
- **Designed for gap analysis** **(more later)**

Categorization of data standards by content

Domain-specific

- By sub-domain: Which domain classification? Attempted one based on FAO + USDA classifications
  How far to go with **neighboring disciplines**?

- By data type:
  alignment with GODAN "Agricultural Sector Package" for the Open Data Charter



Sub-domains

- Agricultural Research, Technology and Engineering
- Agro- Economics, Business and Industry
- Animal Science and Animal Products
- Education and Agricultural Extension
- Farms and Farming Systems
- Fisheries and Aquaculture
- Food and Human Nutrition
- Forest Science and Forest Products
- Government, Agricultural Law and Regulations
- Health and Pathology
- Natural Resources, Earth and Environment
- Plant Science and Plant Products
- Rural and Agricultural Sociology

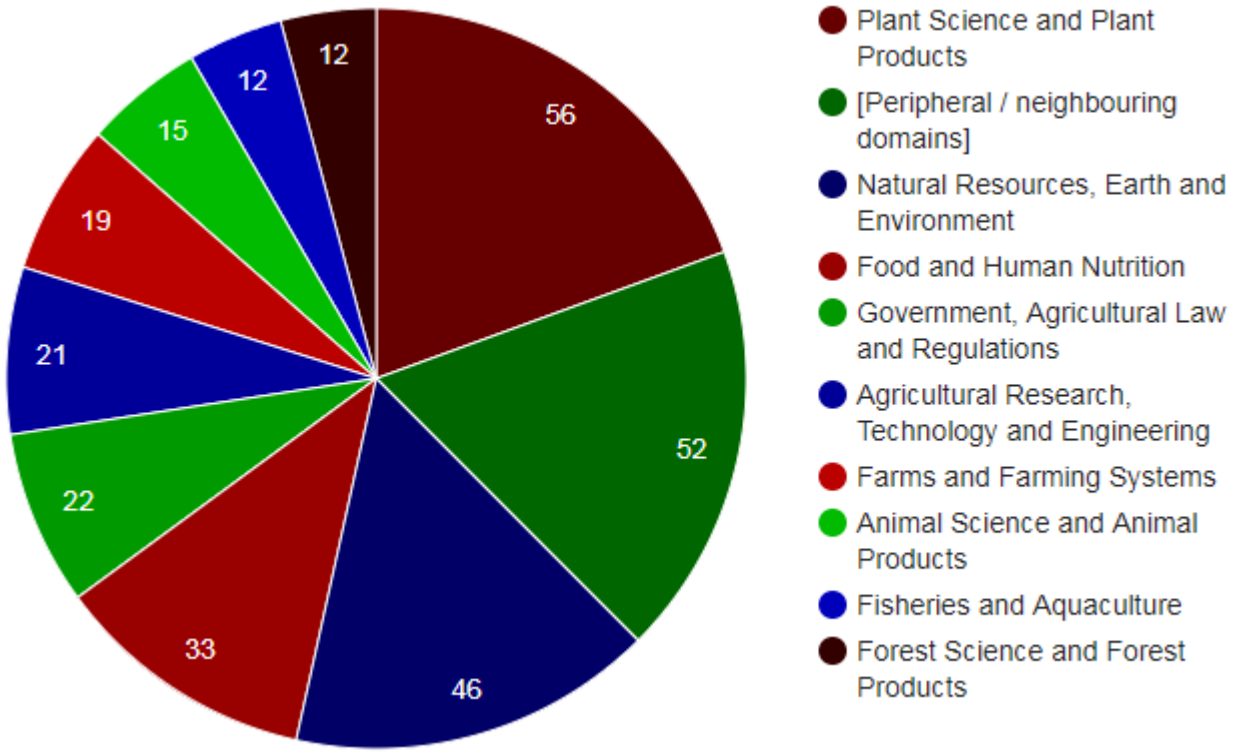
Map of standards so far

Total 375

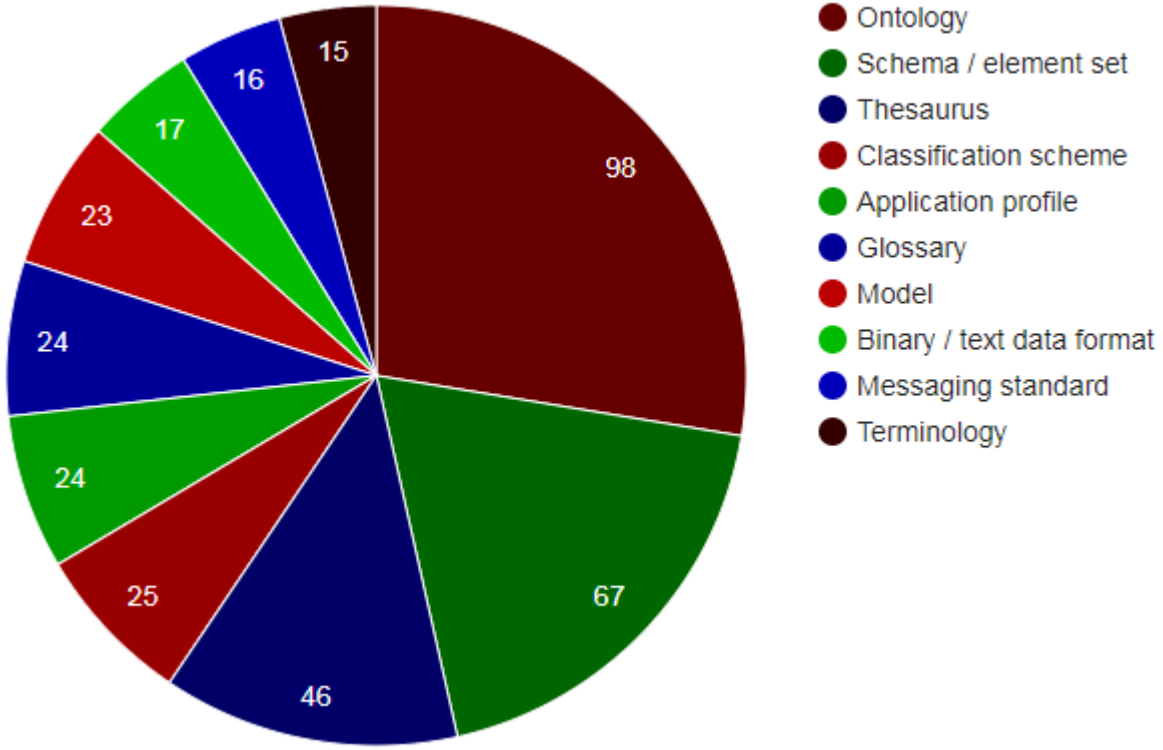
VEST / AgroPortal
MAP OF STANDARDS

vest.agrisemantics.org

Number of data standards by domain



Number of data standards by type



Criteria for gap analysis

To identify gaps, need to identify assessment criteria.
Criteria developed based on:

- The assessment process used by the **UK Government's Open Standards Board**
- The **ODI Open Data Certificates** criteria

Categories of assessment:

Fitness to purpose | **Adoption** | **Usability** | **Openness**

Example of adoption assessment criteria

Used in software	Is this standard used in software tools (e.g. as controlled values for some fields, or as export format, or as data model)?	Yes, in many tools that are very popular	5
		Yes, in a few tools that are very much used	4
		Yes, in a few tools	3
		Yes, in 1-2 tools	2
		No	0
		Not clear - N/A	1
Used in datasets	Is this standard used in datasets (e.g. as serialization format, or as data model / element set, or as controlled values for some columns / dimensions)?	Yes, in many by many providers	5
		Yes, in many, by a few providers	4
		Yes, in a few	3
		No	0
		Not clear - N/A	1
Endorsed	Does the standard have a strong support from different interest groups?	Yes, very strongly	3
		Yes, moderately	2
		No	0
		Not clear - N/A	1
Regulatory	Is the standard published by a recognized standardization body or as a government directive?	Yes	3
		No	0
		Not clear - N/A	1
Long-term, sustainable	Is the maintaining organization a long-standing and authoritative body? Is the maintainer committed to sustain and preserve the standard?	Yes, highly	3
		Yes, reasonably	2
		No	0
		Not clear - N/A	1
Participatory, collaborative	Is participation in the creation process of the standard open to all relevant stakeholders?	Yes	3
		No	0
		Not clear - N/A	1

Example of usability assessment criteria

Versatile	Is the standard available in different formats for different technologies? (e.g. XML, JSON, RDF)?	Yes, many formats	3
		Yes, 2-3 similar formats	2
		No	0
		Not clear - N/A	1
Served by APIs	Are there APIs and web services that allow applications to work with the vocabulary? Choose as many answers as you need.	Yes, to get web- or user- friendly results	3
		Yes, to lookup terms / concepts using several parameters	4
		Yes, to automatically annotate text or data	5
		Yes, to perform cross-walks between vocabularies	6
		Yes, to extract / lookup subsets of vocabularies	7
		No	0
		Not clear - N/A	1
Manageable	Is the standard managed in a collaborative environment? Is it managed on a specialized vocabulary management platform?	Yes, on specialized vocabulary management platform	3
		Yes, on a collaborative environment (e.g. Github)	2
		No	0
		Not clear - N/A	1

Example of openness assessment criteria

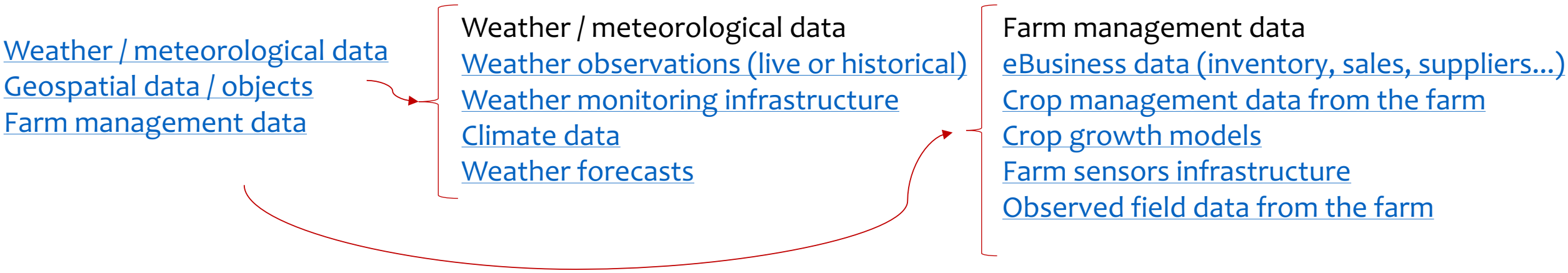
Machine-readable	Is the standard available in machine-readable formats? (XML, CSV, RDF...)	Yes	3
		No	0
		Not clear - N/A	1
Meaningful	If machine-readable, is the standard serialized using the appropriate vocabulary language (RDFs, OWL, SKOS, OBO) and semantically appropriate values?	Yes	3
		No	0
		Not clear - N/A	1
Referenceable	Does it use URIs dereferenceable as URLs as identifiers of classes, properties and instances?	Yes	3
		No	0
		Not clear - N/A	1
Linked	Is the standard available as Linked Data? I.e. serialized as RDF and above all linking to URIs in other vocabularies?	Yes	3
		No	0
		Not clear - N/A	1

Map of standards – topic 1: weather data

- Specific use case of **weather data used in farm management information systems**

- **Coverage**
  - Total of **65 weather and use-case related data standards** http://vest.agrisemantics.org/by-type-of-data/7623+7550+7626/7623/7550/7626

- **Refined classification of weather and use-case data types**



Examples of data standards relevant for topic 1

1 OGC Observations and Measurements - XML Implementation

2 METCE (Modele pour l'Echange de Temps, Climate et Eau)

3 NetCDF Climate and Forecast (CF) metadata conventions

4 BUFR - Binary Universal Form for the Representation of meteorological data

5 GRIB - General Regularly-distributed Information in Binary form

6 Aviation Routine Weather Report (METAR)

7 WMO Codes Registry

8 OGC Sensor Model Language (SensorML)

9 OGC W3C Semantic Sensor Network Ontology

10 Climate Science Modelling Language (CSML)

11 IWXXM (ICAO Meteorological Information Exchange Model)

12 OSCAR Observing Systems Capability Analysis and Review Tool: list of variables

13 CEDA BADC-CSV format

14 ISO 19156:2011 Geographic information - Observations and measurements

15 AgMIP ICASA Master Variable List

16 NetCDF - Network Common Data Form

17 WMO METEOTERM

18 INSPIRE Data Specifications on Atmospheric Conditions and Meteorological Geographical Features

20 HDF5 File Format Specification

21 TMY3 (Typical Meteorological Year)

25 CLIMAT

26 Weather Objects Modelling Language (WOML)

27 ICASA Data Standards for Agricultural Field Experiments and Production

28 OGC Timeseries Profile of Observations and Measurements

29 OGC CF-netCDF3 Data Model Extension standard

30 OGC CF-netCDF 3.0 encoding using GML Coverage Application Schema

31 Common Alerting Protocol (CAP)

Assessment and gap analysis

- Weather data standards
  - Specific use case of weather data used in farm management information systems

Assessment metadata + **consultations with experts**

For geospatial and weather data

- Ben Schaap (GODAN Secretariat)
- Giovanni L'Abate (CRA Italy)
- Simon Cox (CSIRO Australia, RDA)

For weather data for farm management

- Soonho Kim (IFPRI, ICASA standards)
- Andres Ferreyra (AgGateway)
- Hugo Besemer (Wageningen UR)
- Francesco Benincasa (RDA Weather IG, Barcelona Supercomputing Center)
- Allard de Wit (Alterra)
- Christopher Brewster (TNO Netherlands)

Weather data standards – Summary

- **Variety** of data models, data formats and vocabularies that are used to exchange these data
- Some older standards are still very much used, either for **legacy** and compliance reasons (like BUFR or GRIB) or because of long-term practice in research (NetCDF)
- Standardization bodies have worked on **geospatial and observations models** and related schemas (ISO/OGC, especially the ISO19100 series), starting to be used also by the meteorological community (CSML, METCE, IWXXM schemas)
- Recently API-based weather data services started serving **data on demand** and in application-friendly formats like Json and user-friendly formats like CSV
- Work on **variable naming conventions**
- **FMIS**: subset of weather data variables; agreement with weather data providers; work on **variable naming**

Gap analysis on weather data standards

a) *The biggest challenges with weather data are related to issues of data availability, **discoverability**, quality, coverage and documentation.*

b) Data standards are **not well documented**, differences between overlapping standards are not clarified.

c) Both for weather data and for farm management data, standardization of **variable names** across the different communities and even within the same community is an issue; more in general, there are many different **code lists** used by different authorities, with limited alignment.

d) Few data standards published as **machine-readable and linked vocabularies**.

→ *Intermediaries still have to do most of the work converting, processing, re-purposing the data between the different steps in the data value chain.*

Key recommendations for better use and usability of weather data standards

- Address discovery issues relating to weather data (improving use of **discovery metadata** to help catalogue and describe data)
- Improve the **documentation and self-description** of existing data standards (creating developer documentation; publishing existing vocabularies in new ways); offer community support, Q&A services
- Identify 2-3 key **code lists** that should be **published in more linkable, versatile formats**; link key code lists and publish existing **alignments**; provide web services for cross-walks.

Next steps on data standards

- **Data publication online help desk**
  - Q&A service; Facilitate authors/experts on standards to register and provide their help
  - Facilitate data publishers to register and ask for help from the registered authors/experts on standards
- **Same iteration for nutrition data and land data as for weather data**
  - Use cases > survey of standards > gap analysis > recommendations
- **Specifications for standard interoperability services**
  - Based on the recommendations, for all 3 types of data
  - Interoperability specifications that make standards (for all 3 types of data) interoperable and reusable
- **Standard interoperability services for pilot interventions**
  - Based on recommendations and specifications of services
  - Pilots for providing interoperability services for standards identified to use cases from the thematic topics

Useful links

- GODAN: http://godan.info
- GODAN Action map of standards: http://vest.agrisemantics.org
- AgroPortal: http://agroportal.lirmm.fr
- The assessment process used by the UK Government's Open Standards Board: Core questions: https://standards.data.gov.uk/core-assessment-questions
- The ODI Open Data Certificates criteria: https://certificates.theodi.org/en/
- Ag Sector Package: http://agpack.info/
- DC KOS Types vocabulary: http://wiki.dublincore.org/index.php/NKOS_Vocabularies
- Blog post on gap analysis and recommendations: http://www.gfar.net/news/gfar-and-odi-lead-work-gap-analysis-and-recommendations-weather-data-standards

  Gap analysis and recommendations to be published soon.

Weather data standards: survey, gap analysis and recommendations

Thank you

valeria.pesce@fao.org