Physical activity standards research

Background research, intended to inform the creation of new data specifications to help standardise how data is published and shared in the sector.

1.0 Introduction

This report has been created as part of the ODI's collaboration with Sport England, to improve and open up information about physical activity and sporting opportunities across England.

In it we share a summary of our findings to date, and present some recommendations for how we may progress the adoption of standards. We cover the following areas:

- Use cases for physical activity standards, and the different groups involved.
- A review of existing standards around events and activities.
    - Transport mechanisms.
    - Describing events and activities.
    - Registers, taxonomies or hierarchies for different physical activity types.
    - The importance of booking.
- Recommendations for next steps.

We have deliberately limited both the technical level of detail and the level of recommendations that this document presents. Our goal at this stage is to share our research and our current understanding of the core requirements in order to facilitate discussion rather than presenting a finished, work solution. We are publishing this information openly in order to invite comment and contributions, and will then progress towards defined solutions within the OpenActive Community Group.

For further background information around how and why we have got to this point, we'll be releasing some more reports soon.

2.0 The importance of open standards

We have found that data is currently often locked away privately and not shared. It is being kept in structures and formats that don't necessarily match with others' data, and can contain huge inconsistency in data values. If it is published, it can be in ways that make it difficult for others to use - for example as a PDF, or a spreadsheet that someone manually updates.

We're looking to change the sector by addressing some of these blockers through the introduction of standards. Standards can help by:

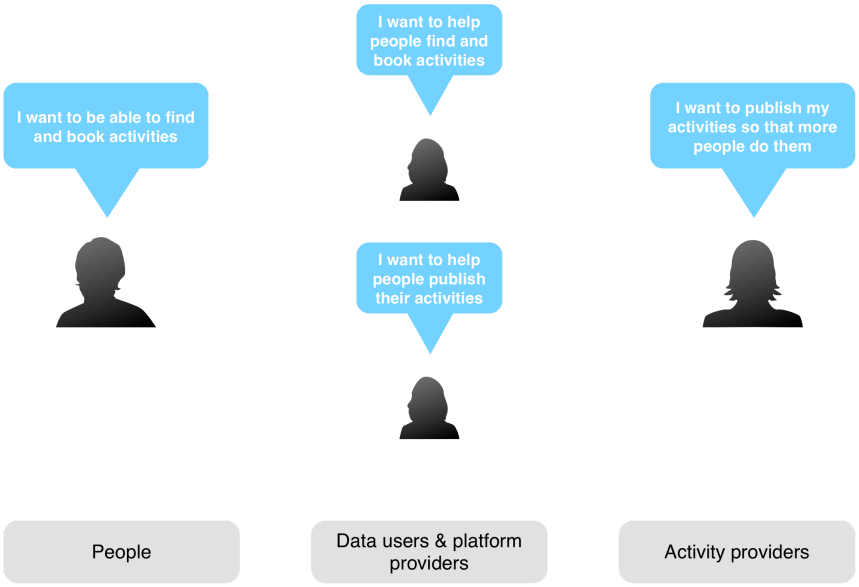
- Providing guidelines for how to make data more useful, and how to publish it - helping those who wish to transform their data collection and publication.
- Ensuring that we're all speaking the same language when it comes to physical activity - if everyone sticks to the same format then we can ensure that there is no misinterpretation, and that all of our data can work well together.
- Making it easier for people to find and use the data that is being published - working together can give greater visibility than working alone.
- Everyone can contribute - we make something that works for everyone rather than being driven by the needs of a few, or by organisations that have an agenda.

Due to the huge variety of needs between the different parties involved in providing opportunity data it is recommended that standards start small and structured, but facilitate inherent flexibility and extensibility. The more situations and scenarios the standards can cover, the more likely they are to be used.

3.0 Understanding the needs around physical activity standards

In order to ensure that we have understood a representative range of challenges facing the sector, our research has considered a number of different user types within the realm of physical activity opportunities.

3.1 What are the different data publisher/user types?



Very simply, we can group different goals as above. In the below table we have aimed to provide slightly more detail around the different publisher/user types as we understand their situations and needs.

User type	Data role	Needs	Constraints	Examples
Activity provider/org with manual/offline record keeping	Publisher	This group may be hesitant to update, and may be initially keen to stay as they are. If they do decide to change, they will likely use other platforms to do so. • Keep digital records in a way that's easy for everyone to use. • Surface opportunities (may not have a website). • Initially combine online and offline management of booking (i.e. won't all be online; legacy behaviour).	• Lack of technical ability • Lack of budget for technology • May not see the benefits	Schools, village tennis courts
Activity provider/org using an off-the shelf publishing/booking system	Publisher	This group is likely already using a solution that they use to manage their opportunity data, possibly as a standalone solution that does not interface with anything else. • Manage activities and opportunities. • Publish these to their own website, social channels, or another platform (i.e. multiple output options required) • Online booking is important and central to their processes.	• Not in control of their own software. • Likely invested with an existing solution. • No in-house resource. • Wants to keep costs low.	Newham & Essex Beagles Athletics club (using opensessions.io)
Activity provider/org using a semi to fully bespoke publishing/booking system that they maintain	Publisher	• Doesn't want too much to change from what they're already doing. • Won't want too much to change in the future, and to keep having to update. • Needs to be informed and involved so that they can plan accordingly, and feel their needs are catered for with standards. • Needs to understand the benefits. • Requires technical references, documentation, support	• Will need convincing that this is a priority to invest in.	GoodGym

Activity provider/org using a semi to fully bespoke booking system that others maintain	Publisher	• May not have the technical skills in house, so are mindful of having to keep paying for changes. • Needs to understand the benefits. • Needs to be informed and involved so that they can plan accordingly, and feel their needs are catered for with standards. • Could be influenced by the service provider making recommendations.	• Mindful of costs • Beholden to the quality, costs and timelines of others.	British Orienteering, Our Parks
Service provider, working with activity providers on their systems	Facilitator	• Requires technical references, documentation, support	• Lack of awareness of open data, and that this is an important issue to push to their clients.	Digital agencies, freelancers
Platform provider - creator of activity management/booking system product	Facilitator	• Life made easier through standardisation of data and formats. • Requires technical references, documentation, support	• Will need convincing that there is demand and that they should invest in offering the functionality. • May have commercial concerns.	BookingBug, OpenTrack, Gladstone, Legend
Data user - aggregator of opportunity data	Reuser	• Life made easier through standardisation of structured data and formats. • More opportunities to use data if it is readable and accessible. • Requires technical references, documentation, support	• Will want to avoid constant changes.	Flexi
Data user - innovator working with sector data	Reuser	• Access to data that is not possible at present. • Requires technical references, documentation, support	• Will want to avoid constant changes.	Startups, search engines 4Global DataHub
People	End-user	• Doesn't care about the underlying transport, formats, etc. Only cares that it's easy to use and gives them what they need.		General public

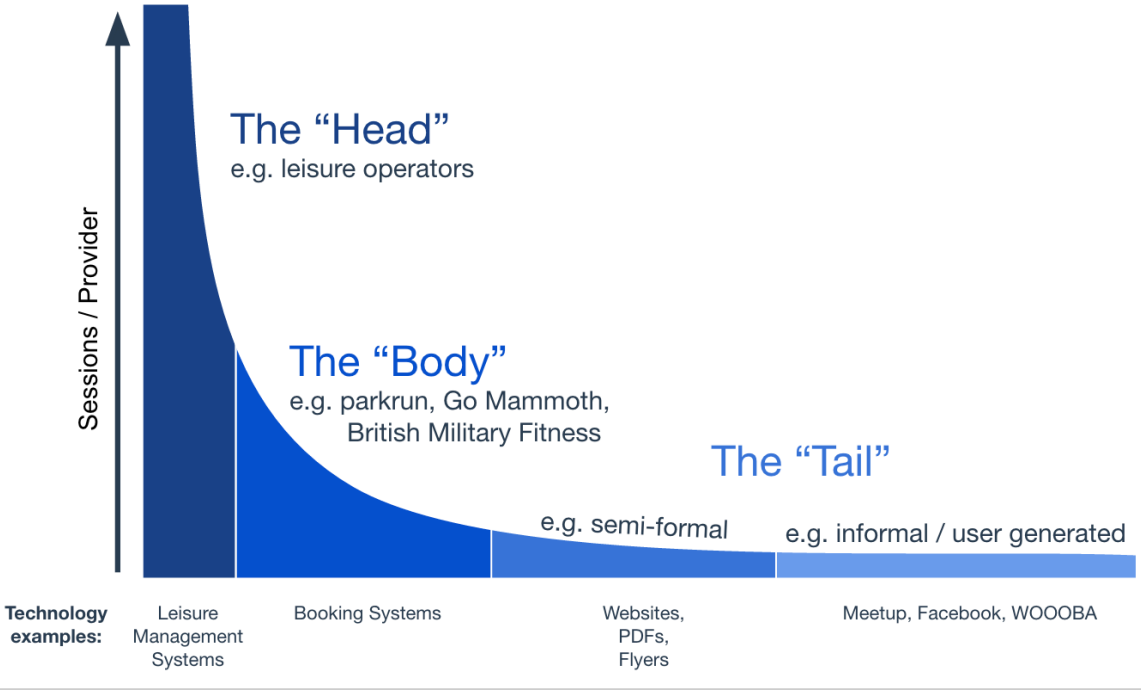
3.2 Current challenges

The sector is reasonably complicated due to the different needs and capabilities from different areas. From our interviews and research, we understand that common challenges to date have included:

Publishers (activity providers)

- Lack of reasons to prioritise publishing data outside of immediate needs.
- Lack of budget available to make technical changes.
- Lack of knowledge around the options available and their impact.
- A perception that there are sector-wide issues with sharing data, and that there is little point trying to improve things alone.

These considerations typically play out to different degrees within different types of organisations or activity providers, which can be viewed as 'the head, body and tail'.



Facilitators and reusers

- Lack of confidence in the reliability and quality of data available.
- Lack of awareness of data that may be available.
- Having to build for a range of variants due to a lack of standards.
- No consistency or processes for managing duplicates - data management overheads.

People

- A reliance on having to call or turn up in person - online data can't always be trusted.
- Fragmented user journeys - getting so far and not being able to book or pay.

- A lack of opportunities available as not everything is online.

3.3 Use cases for standards

Across the different audiences outlined above, it is possible to pull out common use cases in order to view the potential benefits that standards could bring:

Primary use case: The standards allow implementers to publish the activity data in their databases with minimal effort. (The main constraint is that many data providers have limited (and often not continuous) developer resources.)

Use cases for publishers

- Manage the activities and opportunities that they have in-house, using standard fields and values in order to keep the creation process simple and quality high - to improve data quality internally, and to know that it meets industry standards when shared.
- Without the need for an in-house team, enter data about the activities and opportunities they have on a third party platform in a simple way, knowing that it can then be used to publicise the activities more widely.
- Publish activities in a variety of formats, for visibility across platforms (Google, own website, featured on aggregator site etc), with confidence that it will work everywhere.
- Report on their activities and opportunities, and be able to easily combine and compare their data with other data sets.

Use cases for facilitators and reusers

- Create a product for publishers who have no internal development capability, who wish to use a third party to help them publish their data on the web.
- Create an aggregator facility to help people find information about opportunities near them.
- Use sporting data for research and analysis, confident in the quality and the ability to structure queries.
- Understand customer needs and wants against the opportunities available - identify gaps in the market and create new, innovative solutions.

Use cases for people

Due to the increase in services available because of the improved, increased, standardised data, be able to easily find opportunities to meet goals including:

- Get fit/lose weight
- Spend more time outdoors
- Meet people
- Find something for their kids to do

- Etc.

Based on the current conventions, end-user consumers are often provided the means to search by criteria including[1]:

- Location - either near their current location, or near a destination known to them (e.g. their work).
- Date - Either over a broad range (e.g. this weekend), or a specific time (e.g. 6pm tonight).
- Defined categories - e.g. activity types such as Yoga, Cycling, or age groups, disability, gender...
- Tags - more broad options - e.g. relaxing, fun, social...

These criteria may be done individually, or in combination - e.g. What can I do at 6pm near my work? It is recommended that further research is done into the needs of end-users, and used to inform any impact on the standards.

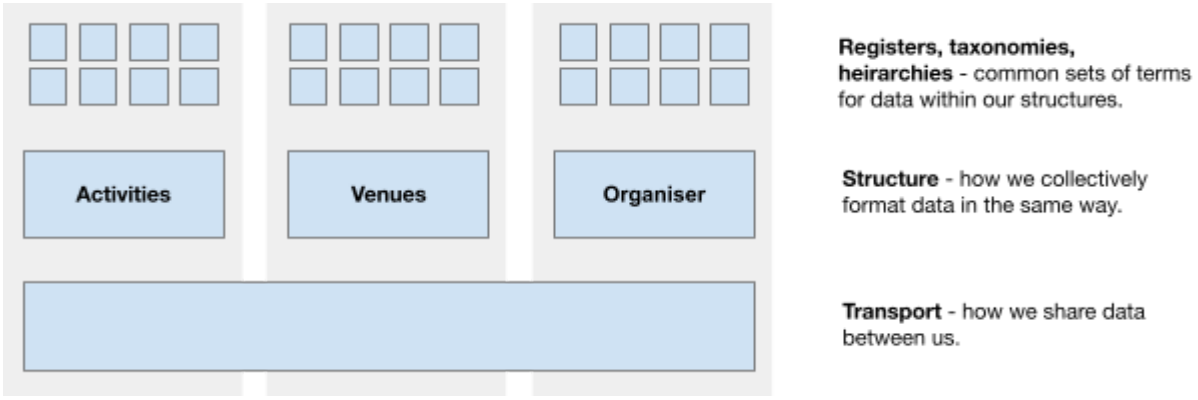
[1] http://www.kaleisure.com/physical-activity/programmes/find-a-physical-activity, http://www.activecumbria.org/developing-sport/physical-activity-search/, http://www.bbc.co.uk/sport/get-inspired, https://beta.getactivelondon.com/, http://londonist.com/category/things-to-do/sport, https://moveyourframe.com/

4.0 Review of existing standards

In order to understand where the gaps are we have conducted a review of existing standards, both from the physical activity sector, and across other notable areas (such as transport). The key topics that we have looked at are transport mechanisms, standard structures for describing events and activities, and registers around activities.

These elements are all related, and can be seen conceptually as follows. The transport mechanism will underlie the data itself, with the structure of entities (such as Activities) being defined by standards, and then being populated with values from standardised registers.

Whilst these 'layers' can be seen as related, they do not necessarily dictate the order in which we need to address standards - i.e. defining transport mechanisms before addressing registers is not necessary. Each can be addressed individually, and in addition each entity can also be addressed individually whilst being aware of the overall goals and strategy.



4.1 Existing transport mechanisms for publishing physical activity data

As the underlying foundation for the recommended standards, the data transport mechanism (the way that data is shared between organisations) is a key piece of the puzzle that we are trying to address.

There are currently many closed, siloed, proprietary data sets that do not facilitate transportation outside of private organisations. If organisations are sharing data, they often deliver it via different methods to one another, privately, with their own standards, meeting their own goals. Some initiatives have sprung up in response to this (e.g. OpenActive), but currently there is no overall consensus.

For our recommendation we aim to draw on wider best practice for data transportation rather than being confined to what is currently the norm within the sector. As such we have reviewed different facets of several data providers' transport mechanisms.

Existing general standards of note

A decision will need to be made regarding the transport mechanism of choice for physical activity data. The following are all individual specifications that may be relevant, but will need to be discussed and tested further against the use cases.

- Atom format (RFC4287 and its update RFC5988) and its publication protocol (RFC5023)
- PubSub
- Activity Streams 2.0
- JSON-LD
- Dat
- GData
- OData

Other publication routes (e.g embedding in HTML pages) may also want to be considered.

In addition to formats, initiatives around API standards, including the Open API initiative, are also relevant for consideration.

Selected existing implementations of note

The following is a summary of a selection of products and services that we have reviewed in order to better understand options and the requirements that we will have around data transport. The publication examples cover both direct data sources and data published by aggregators.

This is by no means intended as an exhaustive review of all of the data available within the sector (and beyond), and we would welcome hearing from organisations around how the following points relate to your data at present, and how they could relate in the future.

Key areas that we focused on included:

- **Discovery** - How does a data user find the latest data from activity providers? Is there a register? Is it discoverable from the home page of a provider?
- **Access** - What is the format for the data from the activity providers? Is there paging through the activities? Are there choices about how the activities are published (eg in JSON and in CSV)?
- **Querying** - Does the implementation support just synchronising/indexing of data or querying?
- **Updates** - How does a data user keep its understanding of what activities a provider is offering up to date? Polling? Subscription?

OpenActive (openactive.io)

Discovery	Discovery is primarily via the openactive.io website, advocacy, and through data set-specific homepages that are created (using GitHub pages) as part of the publishing process. Whilst appearing high up for specific terms such as “sky ride data”, SEO is minimal for more generic terms, suggesting that more could potentially be done in terms of supporting content and linking through to the data homepage. Data sets are also not promoted widely by the individual publishers at present.
Access	Advocates the creation of a RESTful API that outputs JSON, conforming to the Realtime Paged Data Exchange (RPDE) standard defined on the website. CSV/XML are not included in the standard at present. The biggest issue has been the implementation of paging. The implementation intends to be simple but can go very wrong if the concept is not understood. There is a schema validator available, but at present the organisation typically provide manual support to aid implementers. This is unlikely to scale and will require improved tools and documentation.
Querying	No querying is currently offered (outside of the concept of paging).
Updates	The RPDE specification aims to be as simple as possible for publishers, and as such whilst the paging approach will always provide the latest version, the burden on checking for updates is placed on the re-user.
Comments	Standards being championed have focused on the transport mechanisms rather than dictating heavily the format/values of activity data. The format addresses deletion of data through flags in the feed and a paging mechanism. At present because of the size of the data sets being worked with they have not had to clean up deletions, but this may become a consideration after a period of time.

BookingBug (bookingbug.co.uk)

Discovery	Data is an element of the product/platform, and is not shared more broadly. Data is not open at present.
Access	The platform offers embeddable widgets, a RESTful API a JavaScript SDK (runs off the API). Integration with certain other products are available as official features.
Querying	Querying is available.
Updates	Updating data is the responsibility of the reuser.
Comments	• General purpose online booking, appointment, scheduling software. • Tools and support are available. • Documentation uses Swagger.

OpenTrack (opentrack.run)

Discovery	Data is not yet being released, but intentions are stated clearly throughout the site.
Access	To be confirmed - it appears that the intention is to create a RESTful API with OAuth2 and JSON.

Querying	To be confirmed but appears to be planning querying.
Updates	To be confirmed but appears to be planning to require polling.
Comments	• Working within the scope of athletics, but there are large crossovers of intention. • The project makes use of other existing open data sets, and is working to get key public or semi-public data sets released formally as open data.

Bookwhen (bookwhen.com)

Discovery	Whilst the API is documented on the developer site, data is intended to be used by Bookwhen schedule pages rather than being open for use.
Access	Events data is accessed via authentication, with an HTTP endpoint returning JSON. The structure is detailed on the developer site, with no other optional formats available.
Querying	There is an endpoint to get all events, and it is also possible to query for a specific event. No other querying is available.
Updates	Via polling.
Comments	• The results of the call to list all events are limited in terms of fields - more detailed information is only available through the secondary call to retrieve event-specific details.

ACTIVE Access (developer.active.com)

Discovery	The developer documentation and the open data aspect is not obviously promoted on the main active.com website outside of a link to ACTIVE Access in the footer (requires knowledge of what this is). Does not rank highly for terms such as “events and activity data”.
Access	APIs and widgets are available to surface data. API requests are made via HTTP endpoints, and the API supports a variety of response formats (XML, JSON, iCal, RSS), however the options do vary depending on the API used. An API key is required to make calls. Error handling is documented. Widgets provide a form for customisation, generating a collection of links to a stylesheet, several JavaScript files (including an older version of the jQuery library), and a call to render markup. Requests are throttled (2 calls per second) and limited to 500,000 calls per day as standard.
Querying	Querying is possible.
Updates	Via polling.
Comments	• i/o docs section provides a test service query tool.

TransportAPI (transportapi.com)

Discovery	Well publicised, housed within its own website.
Access	Via a RESTful API that serves JSON. Filters and limits are available but paging is not available.

Querying	Querying is available
Updates	Updating data is the responsibility of the reuser.
Comments	• Provides human-readable T&Cs. • Support is available. • Standards are used (e.g. station and bus codes, time duration offsets).

OpenTable (opentable.com)

Discovery	OpenTable have a direct relationship with restaurants (data providers), and control the provision and channels that data can be released through.
Access	Data is accessible via managed website/apps, embeddable widgets, Electronic Reservation Book (ERB) software, or FTP if a member of the affiliate program. There is also an unofficial API which serves JSON/JSONP. This is not authenticated but is throttled by IP.
Querying	Available.
Updates	The third party API places the responsibility on the user. Other mechanisms are obscured.
Comments	A much more closed approach, but notable because of the different options that data is shared by.

Additional suggestions have been made to review APIs from Gladstone, XN, and MINDBODY, but we welcome other information.

4.2 Requirements for future standards around transport mechanisms

With the above in mind, the following will be key requirements to consider during the definition of standards. Our goals will be to trade off between complexity for publishers, and complexity for reusers. Areas to consider around the transport and format of data include:

- How can we get people to make data available easily?
- How can we encourage data being machine-readable?
- How can we encourage data to be real-time?

Requirements include:

- Using an existing mechanism such as HTTP(S).
- Using an existing file format(s).
- To initially focus on the data, not presentation.
- Minimise dependencies.
- Work within mechanisms that the community is already using/adopting so as not to alienate any audiences or reinvent the wheel.
- To include clear T&Cs (human readable - e.g. transport API) & licensing information (including licensing on associated images and other media etc where relevant).
- Providing documentation, support (worked tutorials, sample code, references), and tools (conversion) for the implementation and reuse of data.

Decisions will need to be made around:

- The **format**(s) of the data from activity providers and the subsequent impacts on **machine-readability**, the level of **real-time** data available, **trust**, and **standards adoption**.
    - Whilst static PDF and CSV files may be easier for the 'tail' organisations to generate, these will not be simple for re-users to work with, and will not have the benefit of being real-time data. The concern is that data is already seen as unreliable, and this will likely maintain the perception. In order to benefit the sector, and these standards, a drive for real-time, API-led data is preferable.
- Define the importance of **paging**.
- Whether the initial standard will support indexing of data or full **querying** (introducing a greater level of complexity and order of magnitude in implementation cost).
- Whether the reuser is responsible for polling for **updates**, or whether they are able to subscribe to changes.
- **Authentication/rate limiting** - placing the burden of authentication and rate limiting on the provider is not recommended as mandatory.
    - Ideally data should be unrestricted by default, and then if a CDN/rate limiting/authentication is needed, to introduce this later. This is following the

principle of aiming to keep the publication of data as simple as possible initially.

- **Associated assets** - one area of consideration is if we include the ability to add media/image URLs to activity data. This may lead to hotlinking of the resources, and will require guidance to cache locally or otherwise avoid overloading the original source.
- **Publisher/general dataset information** - whilst defining the standards around activity data itself will be important, the additional metadata and licensing should also be defined within supporting data. This will also aid discovery, and could be used to build credibility through schemes such as Open Data Certificates.

The head and the tail

Whilst considering solutions that will work for the whole sector, it may be that not every potential publisher would need to adhere directly. In order to support the 'tail' organisations it may be that supporting platforms (e.g. third party data collection/publishing options) need to be in place, rather than requiring the 'tail' to develop their own data feeds. Outside of this, better tools and documentation will be needed in order to support the publication of new data sources across all bodies.

4.3 Existing standards for describing events and activities

The key questions we have posed to frame this area of research were "*how do people generally describe **events**, and how do they generally describe **sport**?*".

Selected standards of note within the events/activity landscape generally (including but not limited to sport) have been included below. In general, it has been found that within the realms of activities, there is a large focus on sport.

Sport England

Some work has already been done by Sport England in this area, and whilst it is not exclusively focused on events/activities, we have reviewed a Draft Sport England Taxonomy as part of our research. This includes some data from the Active People dataset, and Sport England also have Active Places Power, which covers locations more than activities, but does include some crossovers. This is very much focused on generally describing activities and locations rather than specific events.

SportsML

SportsML is intended as an industry standard XML format/vocabulary for sports scores, lineups, schedules, standings and statistics. Focuses on sporting events specifically rather than opportunities. This is an open standard, welcoming contributions through a forum. SportsML has been adopted by groups including AP mobile (USA), APA (Austria), BBC (UK), ESPN (USA), NTB (Norway), PA (UK), Univision (USA/Mexico), XML Team Solutions (USA), Yahoo! Sports (USA).

Schema.org

Schema.org has a large amount of crossovers with what we are looking to do. Whether it's the definition of standards around dates, through to events, or venues, many guidelines are already in place and we would look to work with these. Key objects of note include:

- Event (parent of ChildrensEvent, DanceEvent, SportsEvent...)
- Organization (parent of EducationalOrganization, SportsOrganization...)
- Person
- Place (parent of AdministrativeArea (City, Country, State…), Landform (parent of BodyOfWater, Mountain, Waterfall…), CivicStructure (parent of Beach, EventVenue, Park, Playground…), LocalBusiness
- Intangibles including Language, DayOfWeek, GenderType, Geocoordinates, OpeningHoursSpecification, PostalAddress…

When it comes to defining the structure and values for activities, schema.org also provides a standard set of data types (Date, DateTime, Text, URL, Number, Time etc).

W3C Sport Schema Community Group

The Sport Schema Community Group has been working on individual areas, such as scores for head to head events, with a view to feeding these into schema.org. There are currently 14 participants in the group, with discussion being channeled through IRL, a mailing list, arranged calls, and a GitHub repo for feeding back into the core.

Google

Google's Event Markup standard allows for structured information to be surfaced as part of search engine results. The event markup is aligned with schema.org and provides examples around individual events, event listings, and pricing-based queries. Data can be marked up on websites or sourced from elsewhere, and at present is focused on concerts, venues, theatre performances, sports events, comedy, and festivals. As such there is flexibility around dates and other attributes. JSON-LD, Microdata, or RDFa are supported.

BBC Sport Ontology

The BBC Sport ontology originated from a BBC use case, but can be applied more broadly. It is intended as an ontology for publishing data around competitive sports events rather than activities, and references other work such as SportsML.

Centre for Digital Music Event Ontology

An Event Ontology was developed in the Centre for Digital Music at Queen Mary, University of London. The ontology is deliberately simple and flexible, and as such has catered for a variety of event types.

Data shared by activity providers through OpenActive

A number of activity providers have shared data openly through openactive.io, with opportunities ranging from cycling, to table tennis tables, to gym sessions and more. Reviewing this data has shown a range of attributes associated with different activity types. As one example, the Open Sessions fields can be found here.

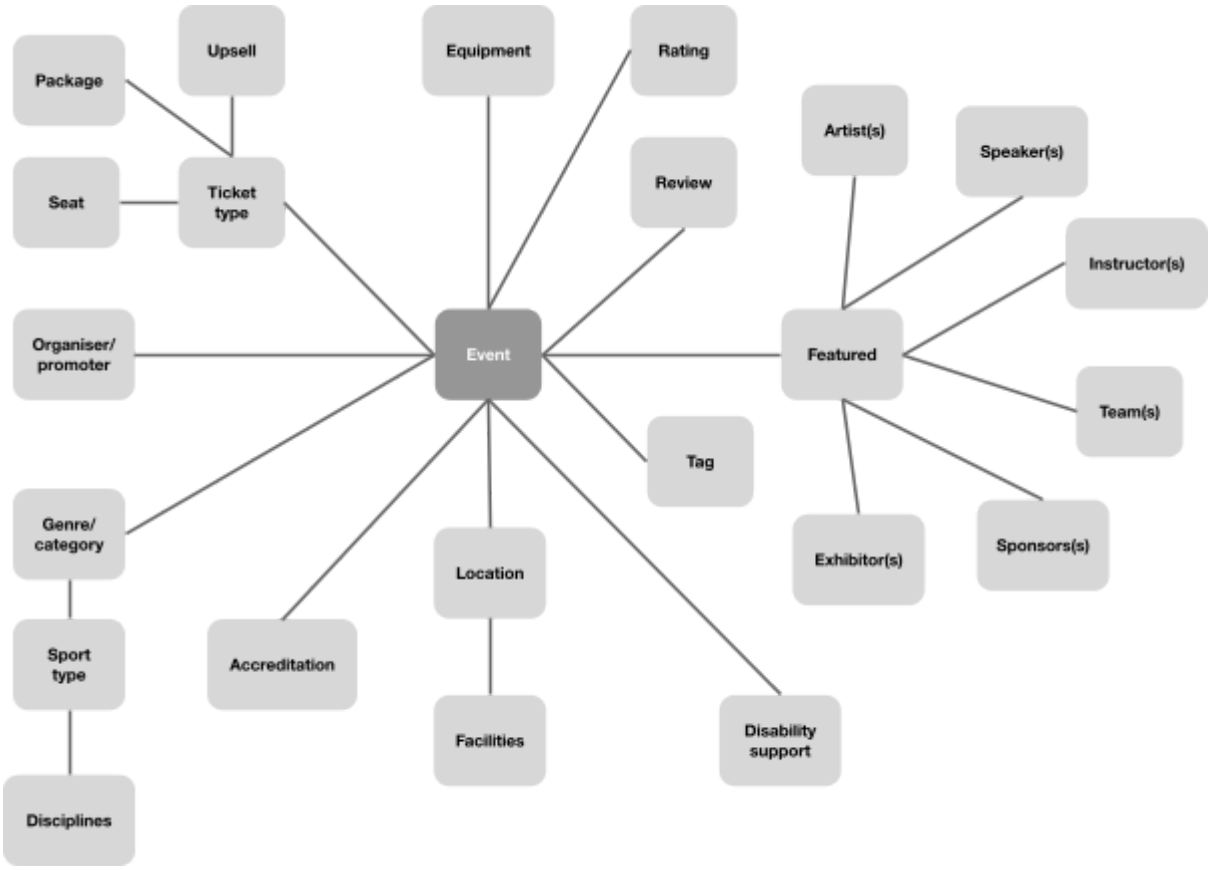
Other examples

Outside of bodies who are looking to define events, sports, and activities, there are many websites that have defined structures and attributes that we can use to expand our understanding of data commonalities, as well as conventions around how users interact with events. More broadly we have looked at sites including:

- Eventbrite
- Ticketmaster
- The Ticket Factory
- TheO2.co.uk
- WembleyStadium.com

Common themes and relationships

From the above, there are key entities which commonly appear throughout different interpretations of events and activities. These are captured in generic form below.



4.4 Requirements for future standards around describing physical activity opportunities

From reviewing the differences in activities (and particularly real-world scenarios of the openactive.io data sets) it is possible to see that the standards will require a balance between concrete requirements, and flexibility to cater for specific needs.

> *"You need some flex to recognise that each sport/activity is unique"*
>
> **League Tennis Association**

Defining physical activities

When looking to standardise a view of a physical activity in terms of the structure, our recommended priorities would be to:

- Promote a simple structure initially, pushing as much complexity as possible outside of the original source of data e.g. tagging, ratings, related events, geographically close events all should be handled by the innovator rather than the data publisher.
- Test the initial small structure to ensure that it is robust, and maintain a well-documented approach for future amendments (whilst maintaining backwards compatibility wherever possible).
- Cater for flexibility alongside the core structure - ensure a 'catch-all' can be used in order to add in data not yet included by the standard, or outliers that will never be covered but which are important to publish.
- Ensure that other standards are made use of where already available (e.g. schema.org).

4.5 Use cases for activities & survey

A set of example activities that can be used to assess whether any data model can represent them, and that can be used as test cases further down the track has been created in the following spreadsheet. These have been informed by real world examples and hypothetical situations.

Activity use cases spreadsheet

In order to reinforce these, and to identify outlier scenarios or gaps that may not have come to light previously, we have put together a survey. This will help us to further understand the type of activities that are being run and the information that is being held. If you'd like to contribute, please fill in the survey.

4.5 Existing registers, taxonomies or hierarchies for different physical activity types

Within the structure of the entities that standards define, we also need to ensure that all parties are working to a common vocabulary and set of options. This is the role of registers.

For our initial research we have focused on sports/activities in line with the views that we heard during our original interviews. Other areas that will subsequently be important will likely cover areas such as venues and equipment.

What is a register?

Registers are an authoritative list of a specific type of thing. This GOV.UK article sets out qualities of a desirable register as being:

- Canonical and having a clear reason for their existence
- Representing a 'minimum viable dataset'
- Live lists, not simply published data
- Use standard names consistently with other registers
- Able to prove integrity of record
- Clearly categorised as open, shared or private
- Containing raw not derived data
- Must have a custodian

What is already being done to define the values for physical activities?

Within taxonomies already available, generally the focus is usually on sports rather than activities for getting active as a whole. Examples of the work being done includes:

- Sport England - Sports that we recognise
- Wikipedia - List of sports
- Active Lives/Active People Survey
- Topend Sports - Complete list of sports from around the world (inc. 200+ with recognised governing bodies)
- Olympic sports (Olympic events represent a small subset of all activities, but are useful for terminology and comparison).
- Work by NGBs and organisations focused on specific areas within the sector - e.g. OpenTrack's work on Athletics.

Whilst venues have not been focused on specifically during our research to date, the Active Places Power open data set is also worth mentioning not only because of its open nature, but as it makes use of individual sports in order to categorise venues.

Hierarchies

These examples are typically semi-hierarchical (nested to a small degree), and provide only top level sporting information rather than additional metadata (difficulty, fitness levels required, etc).

An important element of deciding on standardised values will be to consult individuals and reusers around their needs and expectations for terminology and relationships. These user needs should be aligned with the terminology decided on, but it is important to note that there will be language and differences here - there will not be a single consensus.

Format

The data above varies from being available only as HTML, to CSV, to eventually a JSON feed.

4.6 Requirements for future standards around a physical activity register/taxonomy

Due to the strong level of existing terminology around sports, it is recommended that this is the starting place for a register of physical activities, but that the register progresses beyond this. From the interviews held, the challenge of trying to do too much at once and failing due to a lack of consensus was vocalised by several parties, and this should be considered when deciding on next steps.

Format consideration

The format and structure of the values for physical activities will need to be decided on. One approach could be to have separate SKOS-represented vocabularies, and defining links/mappings between them in order to make it easier for providers or data users to map to whatever is used as the main standard.

Starting data set

From our initial exploration of Sport England's physical activities Ontology, we suspect that this will consist of >350 records (excluding variants). Instead of trying to gain consensus of everything at once, it is recommended that the register starts small, focusing on a segment (i.e. individual sports) and gaining agreement on these individually rather than viewing the register as a whole right from the start. As a starting proposal, it may be possible for each NGB to take responsibility for their own section, using the above Ontology as a starting point, and gaining amendments and eventual consensus from their members.

> *"Athletics, swimming, cycling, triathlon – if you pick these off you would get huge reach / critical mass"*

Sport and Recreation Alliance

Variant terms

One of the challenges cited during interviews was the problem of data quality, and the challenge of standardisation. Many systems in use offer the ability to use free text, which may lead to a variety of spellings and interpretations around what constitutes an activity. As such, an extensive list of mapping variants will likely be the most effective way of addressing this without requiring all adopters to undertake an upfront, full data cleansing exercise of their core data.

We have spoken to an organisation who have a strong existing data set of variant term mappings, and we are keen to gain permission to reuse this if possible.

"They may need to change the strategy for what they call golf; they have their own naming system. If they have to change those descriptions - this could be very difficult. They have programmes that they know work."
Get Into Golf

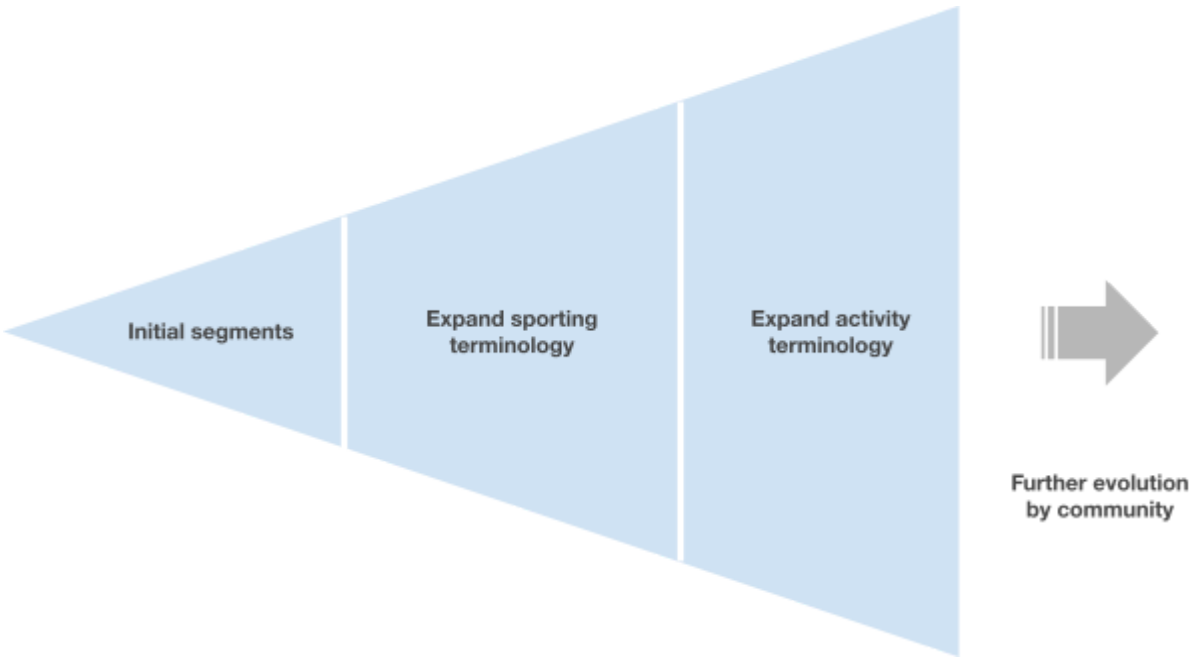
"Adopting external standards (might be the biggest obstacle) . We've had to work hard on getting internal consistency."
National Trust

Further user research and testing

In line with the above, we recommend that further user involvement is sought, both from potential end-consumers and data re-users as well as those who currently hold data. Each activity term may have a slightly different vocabulary based on different user expectations, and these may need to align with the hierarchy and linking in order to support an effective search model.

Evolution over time

Once standards have been adopted, the hope is that over time (with better input standards and eventual data cleanup) the need for variants will lessen, although it is likely that there will still be a need for the most common misspellings and language variants. The data set itself is expected to evolve as new activities emerge, or more organisations are able to contribute.



4.7 Booking

The ability to actually book physical activity opportunities, and for booking to be standardised, has been cited as a hugely important step by many providers and innovators alike. Not only will this enable users to complete their journey, but it will also provide metrics that can be used to measure the impact of publishing data. Whilst this has been heard, due to the complexities involved we are not recommending that the standards community starts with this aspect of the journey.

We feel that without the underlying elements - what activities are, how they're structured, and how they get published - that jumping ahead to booking will not be a success. It's important to lay these foundations first. We're also making this recommendation on the basis of complexity for organisations - booking will require a higher level of integration, including authentication and calls back to internal systems, and this may be too much for certain providers to take on as a first step.

We also feel that there are important conversations to be had by the sector around how to make **open** booking happen. We're hoping that progressing the standards more broadly will aid this conversation once the value of openness has been demonstrated in a more tangible sense.

With all of this in mind, we feel that the roadmap for booking will look something like the following:

1. Define and adopt data **publishing** standards.
2. Define and adopt standards for **reserving places at free opportunities** (may involve the manual sending of information through to full integration with booking systems)

3. Define and adopt standards around **payments** for physical activities involving costs.

5.0 Next steps for the standards

Whilst this document covers our initial recommendations, we want to stress that this is a starting point, and that the success of the standards will be driven by the sector and community itself. As such, the following are suggested next steps, and we'll be forming a proper roadmap within the Community Group.

- Gain feedback on the areas covered in this document.
- Extend research deeper into people's needs, from all perspectives.
- Refine the technical specification for activity data transport, structure, and values.
- Create tools and proof of concepts.
- Improve and extend existing documentation, tutorials, and references.
- Develop a strategy for tracking the impact of opening data/standards.
- Extend standards into other areas - e.g. location/venue data.
- Consider open booking standards - this will be a step where people really start to see the value, but it will be impossible to do this without everything that comes before, and shouldn't be attempted too early.
- Digitisation of the sector needs to happen - bringing people online in the 'tail'.
- Consider applying standards to participation data.

We welcome your comments and contributions, and if you'd like to get involved please visit the OpenActive Community Group.